

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 79.2607 Seconds  
(without alignments)  
388.502 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028

Sequence: 1 TIVSFHONISNNKTERSTA.....LPPDEKFTGAOSGPGQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /SIDSI/gcgdata/genseq/genseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/genseq/genseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/genseq/genseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/genseq/genseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/genseq/genseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/genseq/genseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/genseq/genseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/genseq/genseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/genseq/genseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/genseq/genseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/genseq/genseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/genseq/genseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/genseq/genseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/genseq/genseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/genseq/genseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/genseq/genseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/genseq/genseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/genseq/genseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/genseq/genseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/genseq/genseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/genseq/genseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/genseq/genseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/genseq/genseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/genseq/genseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1028	100.0	1426	23	Human legless homo
2	1028	100.0	1435	22	Human BCL9 homolog
3	277	26.9	140	22	Human polypeptide
4	232.5	22.6	320	23	Mouse beta-catenin
5	232.5	22.6	1494	23	Mouse beta-catenin
6	167.5	16.3	707	22	Human polypeptide
7	162.5	15.8	574	22	WASP homolog prote
8	162.5	15.8	574	22	Amino acid sequenc
9	159	15.5	406	22	Novel human diagno

10	158.5	15.4	256	22	ABUS2945	Human mammary carc
11	158.5	15.4	270	22	ABUS2939	Human mammary carc
12	158.5	15.4	446	22	ABR70063	Drosophila melanog
13	158	15.4	783	13	AAW37151	Mouse neural Wnta*
14	158	15.4	787	13	AAW37152	Mouse neural Wnta*
15	158	15.4	802	22	AAW37153	Mouse neural Wnta*
16	158	15.4	802	22	AAW37153	Mouse neural Wnta*
17	157	15.3	598	22	ABG14000	Novel human diagno
18	155	15.1	293	22	ABUS2937	Human mammary carc
19	155	15.1	572	18	AAW31855	Mycobacterium tube
20	155	15.1	763	18	AAW31852	Protein encoded by
21	154.5	15.0	731	22	ABR74209	Herbicide daily activ
22	154.5	15.0	731	22	ABR74209	Herbicide daily activ
23	153.5	14.9	177	22	ABUS2946	Human mammary carc
24	151.5	14.7	441	22	ABW11413	Human H-2RIIBP. H
25	151	14.7	693	23	ABR69529	Human polypeptide
26	150.5	14.6	533	16	AAW72483	Retinoid C recepto
27	150.5	14.6	533	23	ABR79960	Human retinoid X r
28	150.5	14.6	533	23	AAO19271	Human retinoid X r
29	150.5	14.6	533	23	AAO19282	Human retinoid X r
30	150.5	14.6	533	23	AAO19283	Human retinoid X r
31	150.5	14.6	533	23	AAO19284	Human retinoid X r
32	150.5	14.6	533	23	AAO19285	Human retinoid X r
33	150.5	14.6	533	23	AAO19286	Human retinoid X r
34	150.5	14.6	533	23	AAO19287	Human retinoid X r
35	150.5	14.6	533	23	AAO19288	Human retinoid X r
36	150.5	14.6	533	23	AAO19289	Human retinoid X r
37	150.5	14.6	533	23	AAO19291	Human retinoid X r
38	150.5	14.6	533	23	AAO19292	Human retinoid X r
39	150.5	14.6	533	23	AAO19293	Human retinoid X r
40	150.5	14.6	533	23	AAO19294	Human retinoid X r
41	150.5	14.6	533	23	AAO19295	Human retinoid X r
42	150.5	14.6	533	23	AAO19296	Human retinoid X r
43	150.5	14.6	533	23	AAO19297	Human retinoid X r
44	150.5	14.6	533	23	AAO19298	Human retinoid X r
45	150.5	14.6	533	23	AAO19299	Human retinoid X r

## ALIGNMENTS

### RESULT 1

AAW71229  
ID AAW71229 standard; Protein; 1426 AA.

XX AC AAW71229;

XX XX 18-NOV-2002 (first entry)

XX XX Human legless homologue lgs/bcl9 protein.

XX XX Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;

XX XX tissue proliferation; tumour; cytosolic; cellular disorder; colon;

XX XX blood disorder; cancer; breast; head and neck cancer; brain; thyroid;

XX XX medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX OS Homo sapiens.

XX XX US2002086986-A1.

XX PD 04-JUL-2002.

XX PF 27-JUL-2001; 2001US-0915543.

XX XX 28-JUL-2000; 2000US-221502P.

XX PA (BASL/) BASLER K.

XX PA (BRUN/) BRUNNER E.

XX PA (PROE/) FROESCH E.

XX PA (KRAM/) KRAMPS T.

XX PA (PETE/) PETER O.

XX PI Basler K, Brunner E, Froesch E, Kramps T, Peter O;

XX WPI; 2002-635689/69.  
 DR N-PSDB; AAF88467.  
 XX  
 XX Novel polypeptide useful in therapeutic method for treating disorders  
 PT of cell fate such as cell differentiation or cell proliferation -  
 XX  
 XX Example II; Fig 8B; 41pp; English.  
 XX  
 XX This invention describes a novel polypeptide sharing one or more  
 CC homologous amino acid domains with the legless (lgs) protein, a  
 CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
 CC involved in the formation and maintenance of spatial arrangements  
 CC and proliferation of tissues during development, and in the formation  
 CC and growth of many human tumours. The products of the invention have  
 CC cyostatic activity and can be used to treat cellular disorders, blood  
 CC disorders and cancers caused by over-stimulation of the Wnt pathway,  
 CC where the cancerous condition is colon, breast, head and neck, brain,  
 CC thyroid, medulloblastoma or skin cancer. The product could also be used  
 CC to promote tissue regeneration and repair. This sequence represents the  
 CC human legless (lgs) protein homologue lgs/bcl9 described in the  
 CC disclosure of the invention.  
 XX  
 XX Sequence 1426 AA;  
 XX  
 XX Query Match 100.0%; Score 1028; DB 23; Length 1426;  
 XX Best Local Similarity 100.0%; Pred. No. 7.5e-65;  
 XX Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 1 TIVSFHIONSNKKTSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTLOPTP 60  
 XX 199 TIVSFHIONSNKKTSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTLOPTP 258  
 XX  
 XX 61 PIPAPAPKAPPRPLDRESGVENKLIPIVSGSPASSTPLPDGCTGNSTPNNAVTPVS 120  
 XX 259 PIPAPAPKAPPRPLDRESGVENKLIPIVSGSPASSTPLPDGCTGNSTPNNAVTPVS 318  
 XX  
 XX 121 QCSNSSADKAPPPVSSCEPTLGNPGLSQOLSHERSLTQIRQIMLFPOBK 180  
 XX 319 QCSNSSADKAPPPVSSCEPTLGNPGLSQOLSHERSLTQIRQIMLFPOBK 378  
 XX  
 XX 181 EFTGAQSGGQQNP 194  
 XX 379 EFTGAQSGGQQNP 392  
 XX  
 XX RESULT 2  
 XX ABBI1808  
 XX ID ABBI1808 standard; peptide; 1435 AA.  
 XX AC ABBI1808;  
 XX  
 XX 11-JAN-2002 (first entry)  
 XX  
 XX Human BCL9 homologue, SEQ ID NO:2178.  
 XX  
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 XX haematopoiesis regulation; tissue growth; immunomodulator; activin;  
 XX inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;  
 XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 XX chronic inflammatory condition; proliferative retinopathy;  
 XX atherosclerosis; coronary heart disease; arterial ischaemia;  
 XX bone disorder; osteoporosis; vascular growth disorder;  
 XX tissue regeneration; wound healing; infection; immune disorder;  
 XX cell culture; drug screening; gene therapy; antiinflammatory;  
 XX antiasthmatic; antiarthritis; haemostatic; antiarteriosclerotic;  
 XX cytostatic; osteoplastic; vasotropic; cardiac; virucide; antibacterial;  
 XX antifungal; vulnery; antiulcer.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200157180-A2.

XX 09-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US01800.  
 XX 03-FEB-2000; 2000US-0496914.  
 XX 27-APR-2000; 2000US-0560875.  
 XX (HYSE-) HYSEQ INC.  
 XX  
 XX Tang YT, Liu C, Drmanac RT;  
 XX WPI; 2001-457740/49.  
 XX N-PSDB; ABA09052.  
 XX  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 XX treating or ameliorating a medical condition in a mammalian subject  
 XX e.g. arthritis and cancer -  
 XX  
 XX Claim 20; Page 256-257; 1963pp; English.  
 XX  
 XX Sequences ABBI0981-ABBI12330 represent 1350 novel human polypeptides, and  
 XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 XX invention also relates to vectors and recombinant host cells comprising a  
 XX nucleotide of the invention, methods of producing the novel polypeptides,  
 XX antibodies against the polypeptides, methods of detecting the nucleotides  
 XX or polypeptides in a sample, and methods of identifying compounds which  
 XX bind to polypeptides of the invention. Although novel, many of the  
 XX polypeptides of the invention have homology to known proteins, thereby  
 XX giving an insight into their probable biological activities, and hence  
 XX potential therapeutic applications. The polypeptides of the invention may  
 XX have various activities including cytokine, cell proliferation or cell  
 XX differentiation activities; stem cell growth factor activity;  
 XX haematopoiesis regulatory activity; tissue growth activity;  
 XX immunomodulatory activity; activin, or inhibin-related activities;  
 XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 XX thrombolytic activities; receptor or ligand activities; or may be  
 XX involved in oncogenesis, cancer cell proliferation or metastasis  
 XX depending on their biological activities. Polypeptides and nucleotides of  
 XX the invention are useful for preventing, treating or ameliorating medical  
 XX conditions, e.g., by protein or gene therapy. Such conditions include  
 XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 XX proliferative retinopathy. Polypeptides involved with tissue regeneration and  
 XX repair (or nucleic acids encoding them) may be used to promote wound  
 XX healing (e.g., of burns, incisions and ulcers), while those with  
 XX immunomodulatory activities may be used in the treatment of viral,  
 XX bacterial and fungal infections in addition to immune disorders.  
 XX Polypeptides with growth factor activity may be used in cell cultures to  
 XX promote cell growth. For example, such polypeptides may be used to  
 XX manipulate stem cells in culture to give rise to neuroepithelial cells  
 XX that can be used to augment or replace cells damaged by illness,  
 XX autoimmune disease or accidental damage. The polypeptides and nucleotides  
 XX may also be used in the diagnosis of the above conditions, and in drug  
 XX screening techniques. The present sequence represents a novel human  
 XX polypeptide of the invention.  
 XX  
 XX Sequence 1435 AA;  
 XX  
 XX Query Match 100.0%; Score 1028; DB 22; Length 1435;  
 XX Best Local Similarity 100.0%; Pred. No. 7.6e-65;  
 XX Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 XX 1 TIVSFHIONSNKKTSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTLOPTP 60  
 XX 239 TIVSFHIONSNKKTSTAPLNTQISALRNDPKPLPQQPPAPANQDNSSQNTLOPTP 298  
 XX  
 XX 61 PIPAPAPKAPPRPLDRESGVENKLIPIVSGSPASSTPLPDGCTGNSTPNNAVTPVS 120  
 XX 299 PIPAPAPKAPPRPLDRESGVENKLIPIVSGSPASSTPLPDGCTGNSTPNNAVTPVS 358

QY 121 QGSSSSADPKAPPPVSSGPPPTLGENPGLSQLEHRSLSQTLRDQIORMLPDEK 180  
 DB 359 QGSSSSADPKAPPPVSSGPPPTLGENPGLSQLEHRSLSQTLRDQIORMLPDEK 418

QY 181 EFTGAQSGGQPNP 194  
 DB 419 EFTGAQSGGQPNP 432

## RESULT 3

AA05855  
 ID AAC05855 standard; Protein; 140 AA.

AC AAC05855;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 19747.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

PN WC200164835-A2.

PN 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-514838/56.

DR N-PSDB; A185786.

XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -

XX Claim 20; SEQ ID NO 19747; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (A179941-A193841) and  
 CC the encoded proteins (AA00010-A01910) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 140 AA;

Query Match 26.9%; Score 277; DB 22; Length 140;

Best Local Similarity 91.4%; Pred. No. 1.7e-12;

Matches 53; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 TIVSFHIONISNKKTERSTAPLNTQISALRNDPKLPQPPAPANQONSQNTLRQP 58

DB 81 TWVSFHIONISNKKTERSTAPLNTQISALRNDPKLPQPPAPASQDQNSQNTLRQP 138

## RESULT 4

AAU78461  
 ID AAU78461 standard; Protein; 320 AA.

XX AC AAU78461;

DT 02-JUL-2002 (first entry)

DE Mouse beta-catenin nuclear localised protein #2.

XX Mouse; beta-catenin nuclear localised protein; cancer;  
 KW gene therapy; EST; expressed sequence tag.

OS Mus musculus.

PN WO200224738-A1.

PD 28-MAR-2002.

XX 19-SEP-2001; 2001WO-JP08140.

XX 22-SEP-2000; 2000JP-0287876.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PI Akiyama T, Adachi S;

WPI; 2002-330014/36.

N-PSDB; ABK47632.

XX New beta-catenin nuclear localised protein for diagnosis and treatment  
 PT of diseases associated with nuclear localisation of beta-catenin e.g.  
 PT cancer -

XX Claim 2; Page 91-92; 113pp; Japanese.

XX The invention relates to a beta-catenin nuclear localised protein  
 CC and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC mouse beta-catenin nuclear localised protein #2.

XX Sequence 320 AA;

Query Match 22.6%; Score 232.5; DB 23; Length 320;

Best Local Similarity 35.1%; Pred. No. 5.7e-09;

Matches 71; Conservative 23; Mismatches 75; Indels 33; Gaps 9;

QY 1 TIVSFHIONISNKKTERSTAPLNTQISALRNDPKLPQPPAPANQONSQNTLRQP 60

DB 16 SILAHQONVPRAKDQA-----PKVPTPEPLFN---TFSAGTQSQPP 58

QY 61 PIPAPAPKAPRPLDRSP--GVENKLIP--SVGSPASSTPLPDGTGNSPTNNRAVT 117

DB 59 PLPFPFPAPGAPALPPEGPEDTSQDLAPNSVG--AASTGGTGGTHTPTAATANN 116

QY 118 FVSSQSSSSADPKA--PPPPVSSGPEPTLGENPGLSQLEHRSLSQTLRDQIORML 175

DB 117 PLPPGGDPGAPGALLGEATPTGNGQRLVGS--EGLSKQLEHRSLSQTLRDQIORML 174

QY 176 F--PDEKEFT-----GAQSGGP 190

DB 175 LRSGETEFLKPGPGGAGEGGP 196

## RESULT 5

AAU78460  
 ID AAU78460 standard; Protein; 1494 AA.

XX AC AAU78460;





ID AAMS2322 standard; Protein; 574 AA.  
 AC AAMS2322;  
 XX  
 DT 18-JAN-2002 (first entry)  
 DE WASP homolog protein.  
 DE WASP homolog protein.  
 XX  
 KW Actin polymerisation; Ena/VASP; vasodilator-stimulated phosphoprotein;  
 KW metastatic cancer; parasitic infection; cytotoxic; WASP.  
 XX  
 OS Schizosaccharomyces pombe.  
 XX  
 PN WO200171356-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-FR00843.  
 XX  
 PR 22-MAR-2000; 2000FR-0003637.  
 XX  
 PA (CNRS) CENT NAT RECH SCI.  
 XX  
 PA (CURRI) INST CURIE.  
 XX  
 PI Fradelizi J, Friederich E, Golsteyn RM, Louvard D, Noireaux V;  
 PI Sykes C;  
 XX  
 DR WPI; 2001-639148/73.  
 XX  
 DR Identifying modulators of actin polymerization, potentially useful for  
 PT treating tumor metastasis and parasitic infection, using proteins that  
 PT contain Ena/VASP binding sites -  
 XX  
 PS Claim 13; Pages 107-109; 109pp; French.  
 XX  
 CC The present invention relates to a method for identifying modulators of  
 CC actin polymerisation. The method involves using proteins that contain at  
 CC least one binding motif for proteins of the Ena/VASP  
 CC (vasodilator-stimulated phosphoprotein) family in the preparation of  
 CC reagents for identification/screening of molecules that modulate  
 CC formation of the actin cytoskeleton. The proteins used in the method  
 CC (i.e. the proteins with binding motifs) for Ena/VASP proteins) do not  
 CC bind to the Arp2/3 protein complex. The modulators identified by the  
 CC method are potentially useful for treating disorders of actin  
 CC polymerisation, e.g. the metastatic cancer or parasitic infection, and as  
 CC cytotoxic agents. The present sequence one such protein with binding  
 CC motif(s) for Ena/VASP proteins, which was used in the method of the  
 CC present invention.  
 XX  
 XX Sequence 574 AA;  
 XX  
 Query Match 15.8%; Score 162.5; DB 22; Length 574;  
 Best Local Similarity 30.9%; Pred. No. 0.00097;  
 Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
 QY 7 IONISNKKTERSTA-----PLNTQISA---LNDPKLPQQPPAPANODQ-----NSSQNT 54  
 DB 275 IAPVSNPAINSTSKPLPPFSRVSAAALANKKRPFPFPPPSRRNRKGPPIGNGSSNS 334  
 QY 55 RLQPTPP-----IPAPAKPAAPRPLDRESGVENKLIIPVSGSPA-SSTPLPDG 104  
 DB 335 SLPPFPFPPRNSAAGSTPLPQGRSAPFPFPPPSRPSAPSTGQRPPLSSSRVSNPAPPPA 394  
 QY 105 TGNSTPNNRVAVTVSGNSSSSADPKAPFPFPPVSSGEPPTL 146  
 DB 395 IPGRSAP---ALPLPLGNASRTST--PVPPTPSLPSAPPSL 431  
 RESULT 8  
 AAG67370  
 ID AAG67370 standard; Protein; 574 AA.  
 XX  
 AC AAG67370;

XX  
 DT 13-NOV-2001 (first entry)  
 XX  
 DE Amino acid sequence of a yeast WASP protein homologue.  
 XX  
 KW Wiskott-Aldrich syndrome protein; WASP; actin cytoskeleton;  
 KW cell motility; actin polymerisation; cancer; parasite infection;  
 KW embryonic development; immune response; wound repair.  
 XX  
 OS Schizosaccharomyces pombe.  
 XX  
 PN WO200144292-A2.  
 XX  
 PD 21-JUN-2001.  
 XX  
 PF 15-DEC-2000; 2000WO-FR03569.  
 XX  
 PR 16-DEC-1999; 99FR-0015900.  
 XX  
 PA (CNRS) CENT NAT RECH SCI.  
 XX  
 PA (CURRI) INST CURIE.  
 XX  
 PI Noireaux V, Prost J, Sykes C, Friederich E, Golsteyn RM;  
 PI Louvard D;  
 XX  
 DR WPI; 2001-536241/59.  
 XX  
 DR N-PSDB; AAH77922.  
 XX  
 DR New fragments of WASP family proteins, useful for detecting and  
 PT identifying modulators of actin cytoskeleton formation, potential  
 PT anticancer and antiparasitic agents -  
 XX  
 PS Claim 14; Fig 8; 162pp; French.  
 XX  
 CC The present sequence represents a WASP (Wiskott-Aldrich syndrome  
 CC protein) homologue. Peptide fragments of WASP-family proteins of  
 CC eukaryotic cells are used to prepare reagents for detecting compounds  
 CC that inhibit or stimulate formation of the actin cytoskeleton, and  
 CC thus inhibit or stimulate cell motility. The peptides are used to  
 CC detect and identify compounds which are potentially useful for treating  
 CC diseases associated with dysfunction of actin polymerisation,  
 CC particularly metastatic cancer and parasite infection; as cytotoxic  
 CC agents for inhibiting/stimulating formation of the actin cytoskeleton  
 CC and for detecting side-effects on actin polymerisation of  
 CC pharmaceuticals. By modulating actin polymerisation, these compounds  
 CC affect cell motility, embryonic development, the immune response and  
 CC wound repair.  
 XX  
 XX Sequence 574 AA;  
 XX  
 Query Match 15.8%; Score 162.5; DB 22; Length 574;  
 Best Local Similarity 30.9%; Pred. No. 0.00097;  
 Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
 QY 7 IONISNKKTERSTA-----PLNTQISA---LNDPKLPQQPPAPANODQ-----NSSQNT 54  
 DB 275 IAPVSNPAINSTSKPLPPFSRVSAAALANKKRPFPFPPPSRRNRKGPPIGNGSSNS 334  
 QY 55 RLQPTPP-----IPAPAKPAAPRPLDRESGVENKLIIPVSGSPA-SSTPLPDG 104  
 DB 335 SLPPFPFPPRNSAAGSTPLPQGRSAPFPFPPPSRPSAPSTGQRPPLSSSRVSNPAPPPA 394  
 QY 105 TGNSTPNNRVAVTVSGNSSSSADPKAPFPFPPVSSGEPPTL 146  
 DB 395 IPGRSAP---ALPLPLGNASRTST--PVPPTPSLPSAPPSL 431  
 RESULT 9  
 ABG27250  
 ID ABG27250 standard; Protein; 406 AA.  
 XX  
 AC ABG27250;  
 XX

DT 18-FEB-2002 (first entry)  
 DE Novel human diagnostic protein #27241.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 PI Dmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-539362/73.  
 DR N-FSDB; AAS91437.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PT  
 XX Claim 20; SEQ ID NO 57609; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG3037 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://ftp.wipo.int/pub/published_pct_sequences).  
 XX  
 SQ Sequence 406 AA;  
 Query Match 15.5%; Score 159; DB 22; Length 406;  
 Best Local Similarity 32.0%; Pred. No. 0.0012;  
 Matches 41; Conservative 15; Mismatches 62; Indels 10; Gaps 4;  
 QY 33 KPLPLQQQPPAPA-----NQDQNSQNTLRQTPPIAPAPKPA-APPRLDRESRGVE 84  
 DB 20 KSLPLSP 79  
 QY 85 N-KLIPSVGSBASSPLPPDGTGPNSTP-NNRATVTSQGSNSADPKAPPPPVSGE 142  
 DB 80 SP 139  
 QY 143 PPTLGENP 150  
 DB 140 PPSPSPPSP 147

RESULT 10  
 ABUS2945  
 ID ABUS2945 standard; Protein; 256 AA.  
 XX  
 AC ABUS2945;  
 XX  
 DT 14-APR-2003 (first entry)  
 XX  
 DE Human mammary carcinoma-derived DKFZphmcfl\_1c23 homologue #15.  
 XX Human; gene therapy; vaccine; disease treatment; detection.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200112659-A2.  
 XX 22-FEB-2001.  
 PD  
 XX 18-AUG-2000; 2000WO-IB01496.  
 XX 18-AUG-1999; 99US-0149499.  
 PR 28-SEP-1999; 99US-0156503.  
 XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
 PA  
 XX Wiemann S;  
 PI  
 XX WPI; 2001-327840/34.  
 DR  
 PT Nucleic acids having the sequences of clones isolated from libraries of  
 PT different human tissues, useful in recombinant DNA methodologies -  
 PT  
 XX Example III; Page 551; 1095pp; English.  
 PS  
 CC This invention describes novel polynucleotides and polypeptides isolated  
 CC from human cDNA libraries which can be used for gene therapy or in  
 CC vaccines. The polynucleotides of the invention and antibodies encoded by  
 CC them may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression. The products of the  
 CC invention may also be used to identify modulators of expression and  
 CC activity and to down regulate expression and activity. The antibodies of  
 CC the invention may also be used as diagnostic agents for detecting the  
 CC presence of polypeptides in samples. This sequence represents a homologue  
 CC of a polypeptide described in the disclosure of the invention.  
 XX  
 SQ Sequence 256 AA;  
 Query Match 15.4%; Score 158.5; DB 22; Length 256;  
 Best Local Similarity 32.0%; Pred. No. 0.00082;  
 Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;  
 QY 14 KTERSTAPLNTQIGALANDPKPLP--QQPPAPANQDQNSQNTLOP----TPTPAPAPK 68  
 DB 46 KSGPPAPVNLPPPEVKSSPPTPVSSPPAPAKSPFPAPMSSPFPPEVKSPFPAPVSS 105  
 QY 69 P-----AAPRLDRESGVENKLIPSVGSBASTPLP----PDGTGPNSTNNRATVPV 119  
 DB 106 PPPPVKSPFPAPVSSPFPVK---PPPPAPVSSPFPFPVKSPFPAPISGPPFPVKSP 162  
 QY 120 SQGSNSADP-KAPPPP-PVSSGEPTLGENP 150  
 DB 163 PPAPVSSPFPFPVKSPFPAPVSSPFPFPVKSPFPFPVKSPFPFPVKSPFP 195

RESULT 11  
 ABUS2939  
 ID ABUS2939 standard; Protein; 270 AA.  
 XX  
 AC ABUS2939;  
 XX  
 DT 14-APR-2003 (first entry)  
 XX  
 DE Human mammary carcinoma-derived DKFZphmcfl\_1c23 homologue #9.

XX Human; gene therapy; vaccine; disease treatment; detection.  
XX Homo sapiens.  
XX WO200112659-A2.  
XX PD 22-FEB-2001.  
XX PF 18-AUG-2000; 2000WO-IB01496.  
XX PR 18-AUG-1999; 99US-0149499.  
XX PR 28-SEP-1999; 99US-0156503.  
XX (GEHU-) GERMAN HUMAN GENOME PROJECT.  
XX PA Miemann S;  
XX PI WPI; 2001-327840/34.  
XX DR Nucleic acids having the sequences of clones isolated from libraries of  
PT different human tissues, useful in recombinant DNA methodologies -  
XX Example III; Page 550; 1095pp; English.  
XX CC This invention describes novel polynucleotides and polypeptides isolated  
CC from human cDNA libraries which can be used for gene therapy or in  
CC vaccines. The polynucleotides of the invention and antibodies encoded by  
CC them may be used in the prevention, diagnosis and treatment of diseases  
CC associated with inappropriate polypeptide expression. The products of the  
CC invention may also be used to identify modulators of expression and  
CC activity and to down regulate expression and activity. The antibodies of  
CC the invention may also be used as diagnostic agents for detecting the  
CC presence of polypeptides in samples. This sequence represents a homologue  
CC of a polypeptide described in the disclosure of the invention.  
XX SQ Sequence 270 AA;  
Query Match 15.4%; Score 158.5; DB 22; Length 270;  
Best Local Similarity 32.0%; Pred. No. 0.00087;  
Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;  
QY 14 KTERSTAPLNTQISALENPKLP--QQPPAPANQDNSSQNTLQP---TPPIAPAPK 68  
DB 79 KSEPPAPVNLPPPEVKSPPFTVSSPPFAPKSSPPFAPMSSPPFVKSPPFPFVSS 138  
QY 69 P-----AAPPRDLRESQVENKLIFSVGSPASSTPLP---PDGTGNSTNNRAVTPV 119  
DB 139 PPPPVKSPFPAPVSSPPFVKS---PPFAPVSSPPFPVKSPPFPFAPISPPFPVKSPP 195  
QY 120 SQGSNSSADP-KAPPPP-FVSGSEPTLGENP 150  
DB 196 PPAPVSSPPFPVKSPPFPAPVSSPPFPKSPPP 228  
RESULT 12  
ABB70063  
ID ABB70063 standard; Protein; 446 AA.  
XX AC ABB70063;  
XX DT 26-MAR-2002 (first entry)  
XX DE Drosophila melanogaster polypeptide SEQ ID NO 36981.  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.  
XX OS Drosophila melanogaster.  
XX PN WO200171042-A2.  
XX PD 27-SEP-2001.  
XX PF 23-MAR-2001; 2001WO-US09231.  
XX PR 23-MAR-2000; 2000US-191637P.  
XX PR 11-JUL-2000; 2000US-0614150.  
XX PA (PEKE ) PE CORP NY.  
XX PI Venter JC, Adams M, Li PWD, Myers EM;  
XX WPI; 2001-656860/75.  
XX DR N-PSDB; ABL14166.  
XX PR New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX PS Disclosure; SEQ ID NO 36981; 21pp + Sequence Listing; English.  
XX CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (AB857737-AB872072),  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 446 AA;  
Query Match 15.4%; Score 158.5; DB 22; Length 446;  
Best Local Similarity 29.8%; Pred. No. 0.0014;  
Matches 42; Conservative 11; Mismatches 64; Indels 25; Gaps 3;  
QY 33 EKPLQQPPAPANQDNSSQNTLQPTTPIAPA--PKPAAPRDLRESQVENKLIPS 90  
DB 97 QTRTPPPPPPP-----QTFEPAPRPSYGPQTQPPRPPPTPSAPAPPPS 144  
QY 91 VG-----SPASSTPLPDGTGNSTNNRAVTPVSGSNSSADPKAPPPPPVS 139  
DB 145 YGPPQTTPPPRPPQPTPSAPAPRPSYGPQTTPPPPPPTPSAPAPRPPQPPAPQ 204  
QY 140 GSEPTLGENPDGLSGLQLEHR 161  
DB 205 PRSPTSPQPGEVLPDQPKR 226  
RESULT 13  
AAW37151  
ID AAW37151 standard; Protein; 783 AA.  
XX AC AAW37151;  
XX DT 06-JUL-1998 (first entry)  
XX DE Mouse neural Mena+ protein.  
XX KW Neural Mena+ protein; mammalian Ena; Enabled protein; Evi protein;  
XX KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;  
XX KW cell growth; cell motility; mouse.  
XX OS Mus musculus.  
XX PH Key Location/Qualifiers  
FT Misc-difference 378 /note= "encoded by GGN"  
XX PN WO9801755-A1.  
XX PD 15-JAN-1998.

```

XX PF 03-JUL-1997; 97WO-US11669.
XX PR 05-JUL-1996; 96US-0675815.
XX PA (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Gertler FS, Niebuhr K, Soriano P, Mehland J;
XX DR MPI; 1998-101197/09.
XX DR N-PSDB; AAV02998.
XX PT - use in control of cytoskeletal dynamic events in normal and
XX PT abnormal cell morphology, adhesion, motility, growth and
XX PT differentiation
XX PS Example 4; Page 58-60; 77pp; English.
XX CC This protein comprises novel murine neural Mena+. Its amino acid
XX CC sequence was deduced from a cDNA clone (see AAV02998) obtained from
XX CC a mouse brain cDNA library. Neural Mena+ contains an exon that
XX CC introduces 244 amino acids between amino acids 238 and 239 of
XX CC mammalian Ena (Mena, see AAW37148). Two other isoforms, neural
XX CC Mena++ (see AAW37152) and neural Mena+++ (see AAW37153), are also
XX CC disclosed. Unlike Mena, neural Mena isoforms exhibit neural
XX CC tissue-specific distribution. Based on the disclosed Mena and Evi
XX CC genes (see also AAV02996-97) and proteins (see also AAW37148-49), a
XX CC variety of methods and compositions are provided for screening,
XX CC isolating and characterising endogenous and exogenous factors,
XX CC drugs and therapeutic agents useful to evaluate and/or control
XX CC cytoskeletal dynamic events involved in normal and abnormal cell
XX CC morphology, adhesion, motility, growth and/or differentiation. A
XX CC method of detecting a modulator of Mena activity/expression is
XX CC claimed.
XX SQ Sequence 783 AA;
Query Match 15.4%; Score 158; DB 19; Length 783;
Best Local Similarity 26.3%; Pred. No. 0.0028;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;
QY 13 NTERSTAPLNT-----QISALRNDP----- 33
DB 329 NNSRFSSEVNTSSQPPAKSCAWPTSNFSLPSPSPIMISSPPKATGRPVLPVCVS 388
QY 34 KPLDQPPAPANQDN-----SSQNTLQPTPIPAAPKPAAPRPL----- 76
DB 389 SPVQPPSPPTAPNGSLDSVTYVSPPTSGPAAPPPPPPPPPPPPPPLPLPLPLAS 448
QY 77 -----DRESGVEKNLI-----PSVGSASS-TLPPD-GTGNSTNNRAV- 116
DB 449 LSHCGSQSPPTPGTPLASTFSSKPSVLPSAGAPASAEPLNPELGSSASEFGLQAS 508
QY 117 -----TPVSGSNSSADPKAPPPPPVSSGE-----PPTIGENP 150
DB 509 QPAESPTQGG--LVLGPPAPPPPPPLPSGPAYASALPPPGPPP 550
RESULT 14
ID AAW37152 standard; Protein: 787 AA.
XX AC AAW37152;
XX DT 06-JUL-1998 (first entry)
XX DE Mouse neural Mena++ protein.
XX KW Neural Mena++ protein; mammalian Ena; Enabled protein; Evi protein;
XX KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;
XX KW cell growth; cell motility; mouse.

```

```

XX OS Mus musculus.
XX PN W09801755-A1.
XX XX 15-JAN-1998.
XX PD 03-JUL-1997; 97WO-US11669.
XX PF 05-JUL-1996; 96US-0675815.
XX PR (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX PI Gertler FS, Niebuhr K, Soriano P, Mehland J;
XX DR MPI; 1998-101197/09.
XX PT Detection of modulators of Mena and Ena-VASP-like genes and proteins
XX PT - use in control of cytoskeletal dynamic events in normal and
XX PT abnormal cell morphology, adhesion, motility, growth and
XX PT differentiation
XX PS Example 4; Page 60-63; 77pp; English.
XX CC This protein comprises novel murine neural Mena+. Its amino acid
XX CC sequence was deduced from a cDNA clone obtained from a mouse brain
XX CC cDNA library. Two other isoforms, neural Mena+ (see AAW37151) and
XX CC neural Mena+++ (see AAW37153), are also disclosed. Unlike mammalian
XX CC Ena (Mena, see AAW37148), neural Mena isoforms exhibit neural
XX CC tissue-specific distribution. Based on the disclosed Mena and Evi
XX CC genes (see also AAV02996-98) and proteins (see also AAW37148-49), a
XX CC variety of methods and compositions are provided for screening,
XX CC isolating and characterising endogenous and exogenous factors,
XX CC drugs and therapeutic agents useful to evaluate and/or control
XX CC cytoskeletal dynamic events involved in normal and abnormal cell
XX CC morphology, adhesion, motility, growth and/or differentiation. A
XX CC method of detecting a modulator of Mena activity/expression is
XX CC claimed.
XX SQ Sequence 787 AA;
Query Match 15.4%; Score 158; DB 19; Length 787;
Best Local Similarity 26.3%; Pred. No. 0.0028;
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;
QY 13 NTERSTAPLNT-----QISALRNDP----- 33
DB 333 NNSRFSSEVNTSSQPPAKSCAWPTSNFSLPSPSPIMISSPPKATGRPVLPVCVS 392
QY 34 KPLDQPPAPANQDN-----SSQNTLQPTPIPAAPKPAAPRPL----- 76
DB 393 SPVQPPSPPTAPNGSLDSVTYVSPPTSGPAAPPPPPPPPPPPPPPLPLPLPLAS 452
QY 77 -----DRESGVEKNLI-----PSVGSASS-TLPPD-GTGNSTNNRAV- 116
DB 453 LSHCGSQSPPTPGTPLASTFSSKPSVLPSAGAPASAEPLNPELGSSASEFGLQAS 512
QY 117 -----TPVSGSNSSADPKAPPPPPVSSGE-----PPTIGENP 150
DB 513 QPAESPTQGG--LVLGPPAPPPPPPLPSGPAYASALPPPGPPP 554
RESULT 15
ID AAW37153
XX AC AAW37153 standard; Protein: 802 AA.
XX AC AAW37153;
XX DT 06-JUL-1998 (first entry)
XX DE Mouse neural Mena+++ protein.

```

KW Neural Mena+++ protein; mammalian Ena, Enabled protein; Evi protein;  
 KW cytoskeleton; cell morphology; cell adhesion; cell differentiation;  
 KW cell growth; cell motility; mouse.

XX Mus musculus.

XX WO9801755-A1.

XX 15-JAN-1998.

XX 03-JUL-1997; 97WO-US11669.

XX 05-JUL-1996; 96US-0575815.

XX (GBFB) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.

XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX Gertler PB, Niebuhr K, Soriano P, Wehland J;

XX WPI; 1998-101197/09.

XX Detection of modulators of Mena and Ena-VASP-like genes and proteins  
 PT - used in control of cytoskeletal dynamic events in normal and  
 PT abnormal cell morphology, adhesion, motility, growth and  
 PT differentiation

XX Example 4; Page 63-65; 77pp; English.

XX This protein comprises novel murine neural Mena+++ . Its amino acid  
 CC sequence was deduced from a cDNA clone obtained from a mouse brain  
 CC cDNA library. Two other isoforms, neural Mena+ (see AAW37151) and  
 CC neural Mena++ (see AAW37152), are also disclosed. Unlike mammalian  
 CC Ena (Mena, see AAW37148), neural Mena isoforms exhibit neural  
 CC tissue-specific distribution. Based on the disclosed Mena and Evi  
 CC genes (see also AAV02996-98) and proteins (see also AAW37148-49), a  
 CC variety of methods and compositions are provided for screening,  
 CC isolating and characterising endogenous and exogenous factors,  
 CC drugs and therapeutic agents useful to evaluate and/or control  
 CC cytoskeletal dynamic events involved in normal and abnormal cell  
 CC morphology, adhesion, motility, growth and/or differentiation. A  
 CC method of detecting a modulator of Mena activity/expression is  
 CC claimed.

XX Sequence 802 AA;

Query Match 15.4%; Score 158; DB 19; Length 802;  
 Best Local Similarity 26.3%; Freq. No. 0.0028;  
 Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;  
 QY 13 NKTERSTAPLNT-----QIGALNDP----- 33  
 DB 348 NNUSRPSFVNTSPQPAKSCAWTNSFGLPSPPIINISSPGKATGPRVLPVCVS 407  
 QY 34 KLPQOPAPANQDN-----SSQNTLQPTTPIAPAPKPAAPRPL----- 76  
 DB 408 SPVQMPSPPTAPNGSLDSVTYPVSPPTSGAAPPPTPPPPPPPLPLPLAS 467  
 QY 77 -----DRESPGVENKLI-----PSVGSPPASS-TPLPFD-GTGFNSTPNNRAV- 116  
 DB 468 LSHCSQASPPPTPLASTPSKPSVLPSPGAGAPASAEPLNPLGDSASSEPLQNAS 527  
 QY 117 ----TFVSQASNSSSGADKAPPPPPVSSGE-----PPTLGENP 150  
 DB 528 QPAESFTPGQ--LVLGPAAPPPPPPLPSPGAYAGALPPPPGPP 569

Search completed: November 13, 2003, 09:25:18  
 Job time : 80.2607 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:19 ; Search time 28.6848 Seconds  
(without alignments)  
286.155 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028  
Sequence: 1 TTTSFHIQINISNNKTERSTA.....LFPDEKEFTGAQSGRQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2\_6/prodata/1/iaa/5A.COMB.pap.\*  
2: /cgn2\_6/prodata/1/iaa/5B.COMB.pap.\*  
3: /cgn2\_6/prodata/1/iaa/6A.COMB.pap.\*  
4: /cgn2\_6/prodata/1/iaa/6B.COMB.pap.\*  
5: /cgn2\_6/prodata/1/iaa/PCUTUS.COMB.pap.\*  
6: /cgn2\_6/prodata/1/iaa/backfiles.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	161.5	15.7	816	4	US-09-266-225D-12
2	150.5	14.6	533	1	US-07-952-800-2
3	149.5	14.5	525	3	US-08-764-870-7
4	149.5	14.5	525	3	US-08-980-115-7
5	143.5	14.0	446	1	US-07-952-800-4
6	140.5	13.7	8991	4	US-08-714-741-32
7	138.5	13.5	297	2	US-08-580-545B-6
8	138.5	13.5	297	3	US-09-262-653A-6
9	137	13.3	456	2	US-08-819-013-1
10	136.5	13.3	1248	4	US-09-252-991A-24873
11	136.5	13.3	1248	3	US-09-080-897-2
12	136.5	13.3	1248	3	US-09-323-735-2
13	135	13.1	332	3	US-08-818-112-53
14	135	13.1	332	4	US-08-818-111-53
15	135	13.1	332	4	US-09-056-556-53
16	135	13.1	332	4	US-09-072-596-53
17	135	13.1	334	6	5202236-3
18	133.5	13.0	941	4	US-07-757-022B-14
19	133.5	13.0	1022	4	US-07-757-022B-84
20	133.5	13.0	1038	4	US-07-757-022B-74
21	133.5	13.0	1049	4	US-07-757-022B-58
22	133.5	13.0	1140	4	US-07-757-022B-104
23	133.5	13.0	1270	4	US-07-757-022B-44
24	133.5	13.0	1311	4	US-07-757-022B-42
25	133.5	13.0	1314	4	US-07-757-022B-142
26	133.5	13.0	1314	4	US-07-757-022B-50
27	133.5	13.0	1320	4	US-07-757-022B-46

28 133.5 13.0 1320 4 US-07-757-022B-60 Sequence 60, Appl  
29 133.5 13.0 1354 4 US-07-757-022B-48 Sequence 48, Appl  
30 133.5 13.0 1361 4 US-07-757-022B-40 Sequence 40, Appl  
31 133.5 13.0 1363 4 US-07-757-022B-52 Sequence 52, Appl  
32 133.5 13.0 1404 4 US-07-757-022B-62 Sequence 2, Appl  
33 133.5 13.0 1404 4 US-07-757-022B-62 Sequence 62, Appl  
34 133 12.9 1315 3 US-08-699-595-3 Sequence 3, Appl  
35 132 12.8 214 1 US-08-217-321-4 Sequence 4, Appl  
36 132 12.8 331 6 5202236-3 Patent No. 5202236  
37 132 12.8 1231 4 US-08-714-741-41 Sequence 41, Appl  
38 131 12.7 2442 4 US-09-514-247A-10 Sequence 10, Appl  
39 130.5 12.7 325 1 US-08-382-184-2 Sequence 2, Appl  
40 130.5 12.7 325 2 US-08-641-356-2 Sequence 2, Appl  
41 130.5 12.7 325 3 US-09-132-528-2 Sequence 2, Appl  
42 130.5 12.7 325 3 US-09-132-528-3 Sequence 3, Appl  
43 130.5 12.7 325 3 US-08-875-494-2 Sequence 2, Appl  
44 130.5 12.7 325 4 US-09-599-366-2 Sequence 2, Appl  
45 130.5 12.7 325 4 US-09-599-366-3 Sequence 3, Appl

## ALIGNMENTS

RESULT 1  
US-09-266-225D-12  
; Sequence 12, Application US/09266225D  
; Patent No. 6573364  
; GENERAL INFORMATION:  
; APPLICANT: Nandabalan, Krishan  
; APPLICANT: Kingsmore, Stephen  
; APPLICANT: Tchernev, Velizar  
; TITLE OF INVENTION: Isolation and Characterization of Hermansky-Pudlak Syndrome (HPS) Protein Complexes and HPS Protein-Interacting Proteins  
; TITLE OF INVENTION: Interacting Proteins  
; FILE REFERENCE: 15966-523  
; CURRENT APPLICATION NUMBER: US/09/266.225D  
; CURRENT FILING DATE: 1999-03-10  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO. 12  
; LENGTH: 816  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-266-225D-12

Query Match 15.7%; Score 161.5; DB 4; Length 816;  
Best Local Similarity 32.1%; Pred. No. 1.6e-05;  
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;  
QY 17 RSTAPLNTVIALNDKPLP-----QQPAPANQDNQSSNTRLOPPTTP 63  
Db 577 RPAALITSVPAPAPATPTPTTVPOTSPPGVPAGPTGQPOSAGTSIGVPDPACPP 636  
QY 64 APAPKPAAPPLDRESFG-----VENKLIFSVSASSTP-----LPDGT 105  
Db 637 GFAPHPFGPGIVPAPAPQIATSTSLAAQSLVPPGLGSGSTPGVLVFPFGLPPDA 696  
QY 106 G-PNG---TPNNRAVTPVSGNSSSADPKAPPPPVSGSPPTLG 147  
Db 697 GGAQSQMSSESDVNLVT--QQLSKSQVEDPL-----PVSFGTPKSG 738

RESULT 2  
US-07-952-800-2  
; Sequence 2, Application US/07952800  
; Patent No. 5403925  
; GENERAL INFORMATION:  
; APPLICANT: OZATO, KIKIO  
; TITLE OF INVENTION: A NEW MEMBER OF THE NUCLEAR HORMONE RECEPTOR SUPERFAMILY AND A CDNA CLONE THEREOF  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND AND TOWNSEND

```

; STREET: 1 MARKET PLAZA, STEUART TOWER, SUITE 2000
; CITY: SAN FRANCISCO
; STATE: CA
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952.800
; FILING DATE: 19920928
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Kenneth A
; REGISTRATION NUMBER: 31,677
; REFERENCE/DOCKET NUMBER: 15280-21-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 533 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-952-800-2

Query Match 14.6%; Score 150.5; DB 1; Length 533;
Best Local Similarity 31.0%; Pred. No. 7.7e-05;
Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;

QY 31 NDPKPLPQ--QPPAPNQDQSSQNTLRQTPPIPA-----PKAAPPRFDRES 81
Db 86 SSPNPLQGVPPSPG-----PPLPSTAPSLGSGGAPPMPMPPLGSPFP 134
QY 82 GVENKLPVSGPSSSTPLPDGTGNSTNNRATVPSQSSNSADPKAP----- 133
Db 135 VISS-----SWGSGFLPPAPPGFSPVSSQINSSTVSLPGGSGPPDKVPVLGVRLH 190
QY 134 -PPPNVSSG 141
Db 191 CFPFGGPG 199

RESULT 3
US-07-952-800-2
; Sequence 7, Application US/08764870
; Patent No. 6236946
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletterick, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Kushner, Peter J
; APPLICANT: Apriletti, James W
; APPLICANT: West, Brian
; APPLICANT: Shiao, Andrew K
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand
; FILE OF INVENTION: Binding Domains
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764.870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008.540
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008.543
; FILING DATE: 13-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/008.606
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Jackie N
; REGISTRATION NUMBER: 35,966
; REFERENCE/DOCKET NUMBER: UCAL-246/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 525 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-764-870-7

Query Match 14.5%; Score 149.5; DB 3; Length 525;
Best Local Similarity 31.0%; Pred. No. 9.1e-05;
Matches 40; Conservative 10; Mismatches 46; Indels 33; Gaps 5;

QY 31 NDPKPLPQ--QPPAPNQDQSSQNTLRQTPPIPA-----PKAAPPRFDRES 81
Db 78 SSPNPLQGVPPSPG-----PPLPSTAPSLGSGGAPPMPMPPLGSPFP 126
QY 82 GVENKLPVSGPSSSTPLPDGTGNSTNNRATVPSQSSNSADPKAP----- 133
Db 127 VISS-----SWGSGFLPPAPPGFSPVSSQINSSTVSLPGGSGPPDKVPVLGVRLH 182
QY 134 -PPPNVSSG 141
Db 183 CFPFGGPG 191

RESULT 4
US-08-980-115-7
; Sequence 7, Application US/08980115
; Patent No. 626622
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S
; APPLICANT: Baxter, John D
; APPLICANT: Fletterick, Robert J
; APPLICANT: Wagner, Richard L
; APPLICANT: Kushner, Peter J
; APPLICANT: Apriletti, James W
; APPLICANT: West, Brian L
; APPLICANT: Shiao, Andrew K
; TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
; FILE OF INVENTION: UCAL-246/02US
; CURRENT APPLICATION NUMBER: US/08/980.115
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 08/764.870
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 60/008.606
; EARLIER FILING DATE: 1995-12-14
; EARLIER APPLICATION NUMBER: 60/008.543
; EARLIER FILING DATE: 1995-12-13
; EARLIER APPLICATION NUMBER: 60/008.540
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 7

```

	Query Match	14.0%	Pred 143.5;	DB 1;	Length 446;
	Best Local Similarity	30.3%	Pred No. 0.00023;		
	Matches	37;	Conservative	12;	Mismatches 44; Indels 29; Gaps 4;
Oy	31 NDPKPLPQ--QFAPANQDNSSNTRLEPTTPAPAKPAKAAPEPPPLRRSGPVENKI I 88	:	:	:	:
DG	9 SSNPVLSQSIRFPSP-----PCGPITTSAPPMPMPPPLGGSPVPIS--- 51	:	:	:	:

	Query Match	13.7%	Score 140.5;	DB 4;	Length 8991;	
	Best Local Similarity	27.4%;	Pred No.	0.014;		
	Matches	43; Conservative	9; Mismatches	62; Indels	43; Gaps	5;
QY	16 ERSTAPLNTAISALND-----PKPLRQPDPAPANODNSSNTRLRQTTP--PFP 63					
	:::::					
DG	5961 EKTATKKAELKEAEDLKXAVDEPTTAPAFAPAFA-----FTPEPAPF 6007					
QY	64 APAPKPAAAPPRLDRESGVENKLIPSGVSPASTFLPDGTGNSTPNNRVAITPVSSQS 123					
DG	6008 AAPAKPAFKPAKPAPKPAPKPAPKPAKPAPKPAKPAPAPAPAKPKP----- 6055					
QY	124 NSSGSADPKAPPPPPPVSCEPPLLGENPD---GLSQE 156					



DB 6056 --APAPAPAPAPKPEKPAKPAKPEKTKGWKQE 6090

# RESULT 7

US-08-580-5458-6  
; Sequence 6, Application US/085805458  
; Patent No. 5932713  
; GENERAL INFORMATION:  
; APPLICANT: Yoshihisa, Kasukabe  
; APPLICANT: Koichi, Fujisawa  
; APPLICANT: Susumu, Nishiguchi  
; APPLICANT: Yoshihiko, Maskawa  
; APPLICANT: Randy, Allen  
; TITLE OF INVENTION: COTTON FIBER TISSUE-SPECIFIC GENES  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 601 Thirteenth Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/580,545B  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bretschneider, Barry E.  
; REGISTRATION NUMBER: 28,055  
; REFERENCE/DOCKET NUMBER: 04473/068001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202/783-5070  
; TELEFAX: 202/783-2331  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 297 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-580-5458-6

Query Match 13.5%; Score 138.5; DB 2; Length 297;  
Best Local Similarity 31.2%; Pred. No. 0.00038;  
Matches 43; Conservative 59; Mismatches 59; Indels 27; Gaps 6;  
QY 15 TERSTAPLNTQISALRNDKPLQPPAPAPANQDSSQNTLQTPPI--PAPAPKPAAP 72  
DB 78 TKATPPYKPTKA-----PTP-PYKFPAPAPTK-----AFTPPYKPAAPPTKAP 124  
QY 73 PRPLDRESGVENKLIISVCSPASSTPLPDGCTGNSSTNNRAVTPVSGNSSSADPKA 132  
DB 125 TFPYKPPAP-----APTAKPTPP-FKPPAPAPPTKAPTPPYKPTTAPAPPVK 172  
QY 133 PPPPVSSGEPPTLGENP 150  
DB 173 AFTPPYMPPTPTKPTF 190

# RESULT 8

US-09-262-653A-6  
; Sequence 6, Application US/09262653A  
; Patent No. 6166294  
; GENERAL INFORMATION:  
; APPLICANT: Yoshihisa, Kasukabe  
; APPLICANT: Koichi, Fujisawa  
; APPLICANT: Susumu, Nishiguchi  
; APPLICANT: Yoshihiko, Maskawa  
; APPLICANT: Randy, Allen

; TITLE OF INVENTION: COTTON FIBER TISSUE-SPECIFIC GENES  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 601 Thirteenth Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20005  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/262,653A  
; FILING DATE:  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bretschneider, Barry E.  
; REGISTRATION NUMBER: 28,055  
; REFERENCE/DOCKET NUMBER: 04473/068001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202/783-5070  
; TELEFAX: 202/783-2331  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 297 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-262-653A-6

Query Match 13.5%; Score 138.5; DB 3; Length 297;  
Best Local Similarity 31.2%; Pred. No. 0.00038;  
Matches 43; Conservative 59; Mismatches 59; Indels 27; Gaps 6;  
QY 15 TERSTAPLNTQISALRNDKPLQPPAPAPANQDSSQNTLQTPPI--PAPAPKPAAP 72  
DB 78 TKATPPYKPTKA-----PTP-PYKFPAPAPTK-----AFTPPYKPAAPPTKAP 124  
QY 73 PRPLDRESGVENKLIISVCSPASSTPLPDGCTGNSSTNNRAVTPVSGNSSSADPKA 132  
DB 125 TFPYKPPAP-----APTAKPTPP-FKPPAPAPPTKAPTPPYKPTTAPAPPVK 172  
QY 133 PPPPVSSGEPPTLGENP 150  
DB 173 AFTPPYMPPTPTKPTF 190

# RESULT 9

US-08-819-013-1  
; Sequence 1, Application US/08819013  
; Patent No. 5994522  
; GENERAL INFORMATION:  
; APPLICANT: Chan, Andrew C.  
; TITLE OF INVENTION: BLNK PROTEINS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert  
; STREET: Four Embarcadero Center, Suite 3400  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States  
; ZIP: 94111-4187  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/819,013  
; FILING DATE: 17-MAR-1997

CLASSIFICATION: 435  
PRIOR APPLICATION NUMBER: US 08/788,322  
FILING DATE: 24-JAN-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Silva, Robin M.  
REGISTRATION NUMBER: 38,304  
REFERENCE/DOCKET NUMBER: A-64383-1/RFT/RMS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO. 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 456 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-819-013-1

Query Match 13.3%; Score 137; DB 2; Length 456;  
Best Local Similarity 25.9%; Pred. No. 0.00083;  
Matches 51; Conservative 26; Mismatches 58; Indels 62; Gaps 10;

QY 10 ISNNKTERSTAPLNTQISALRNDPKLPQQPPAPANQDNSSQNTLQPTTPIPAAPAP 69  
DB 120 IDNRSSQRHSPFS-----KTLPSKSWP-----SEKALSTLFTALQXP 162

QY 70 AAPRP---LDRESFGV-----ENKLIFSFGS---PASSTPLPDGTGPNSTNNRAVT 117  
DB 163 QVPEKPKGLLEADYVVPVEDENYIHFTESSPPPERKAPMVRSTKPNSS-----T 216

QY 118 PVS-----QGSNSGADPKAPP---PPVSSGEPFTL-----GENPDGLSGEQ-- 157  
DB 217 PASPTGTSRNSGAWETKSPFPAPAPLPRAGKFTPLKTPVASQDNASSVCSEKPI 276

QY 158 -----LEHRSRLQT 167  
DB 277 PAERHRGSGHRQEAQVS 293

RESULT 10  
US-09-252-991A-24873  
Sequence 24873, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 10716136 US/09/252,991A  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 24873  
LENGTH: 142  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24873

Query Match 13.3%; Score 136.5; DB 4; Length 142;  
Best Local Similarity 26.8%; Pred. No. 0.00023;  
Matches 41; Conservative 19; Mismatches 36; Indels 57; Gaps 7;

QY 7 IQNINNNKTERSTAPL-----NTQISALRNDPKLPQQPPAPANQDNSSQNTLQPT 59  
DB 25 VQAFCSDSGKPGTLLNGLFTNAQLSLQPFCKEVE----- 62

QY 60 PPIPAAPKPAAPPLDRESFGVENKLIFSFGSPASSTPLPDGTGPNSTNNRAVT 119

Db 63 PPVPTSPSPSPPP-PPVPSSTPVP-----PCVPTSPSPSPPPPP-SPVPSPP----- 106  
QY 120 SQGNSSSADPKAPPVPSVSSGSPPTLGENPDG 152  
Db 107 -----PPVPSPPSPS-PPVPSPPSPG 127

RESULT 11  
US-09-080-897-2  
Sequence 2, Application US/09080897  
Patent No. 5985574  
GENERAL INFORMATION:  
APPLICANT: King, Mary-Claire  
APPLICANT: Lynch, Eric D.  
APPLICANT: Lee, Ming  
APPLICANT: Morrow, Jan B.  
APPLICANT: Welch, Firi L.  
APPLICANT: Leon, Pedro E.  
TITLE OF INVENTION: Modulators of Actin  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
STREET: 75 DENISE DRIVE  
CITY: HILLSBOROUGH  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94010  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/080,897  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: OSMAN, RICHARD A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UM97-001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 343-4341  
TELEFAX: (650) 343-4342  
INFORMATION FOR SEQ ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1248 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-080-897-2

Query Match 13.3%; Score 136.5; DB 2; Length 1248;  
Best Local Similarity 32.3%; Pred. No. 0.0029;  
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQQ-----PPAPANQDNSSQNTLQPTTPIPAAPKPA-----APRPLDRES----- 80  
DB 578 PLFGDSGTIIFPPAPGD-----STTPPPPPPPPPPPPPPLPGLGTALSPPLPGLSDATIPPP 633

QY 81 -FGVENKLIFSFGSPASSTPLPDGTGPNSTNNRAVTQSGNSSSADPKAPPVPSV 139  
DB 634 PELPVGIVGIPSPSSLPGLGTALP-----PPLPGSARIPPPPPPLPGSAGIPPPPLPGE 689

QY 140 SGEPTLGENPDG 152  
DB 690 AGMPPPPPPLPGS 702

RESULT 12  
US-09-323-735-2  
Sequence 2, Application US/09323735  
Patent No. 6197932

```

; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan B.
; APPLICANT: Welch, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/323,735
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/080,897
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1248 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-323-735-2

```

```

Query Match 13.3%; Score 136.5; DB 3; Length 1248;
Best Local Similarity 32.3%; Pred. No. 0.0029;
Matches 43; Conservative 8; Mismatches 59; Indels 23; Gaps 5;

QY 35 PLPQ-----PPANQNSQNTLQTPPIAPAPKA-----APRLDRES----- 80
DB 578 PLFGSGTILPPFAGD-----STPPPPPPPPPPPLPGGTATSPPLSGDATIPFP 633
QY 81 -PGVKNLIPSGSPASSTPLPDGTGNTNNRAVTVSGNSSSADPKAPPPPPVS 139
DB 634 PLPLPGVGIESPSSLPFGTAIPP-----PPPLFGSARIPPPPPPLPGSAGIPPPPPPLFGE 689
QY 140 SCEPTTGENPDG 152
DB 690 AGMPPPPPPLPGG 702

```

```

RESULT 13
US-08-818-112-53
; Sequence 53, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 53:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-53

Query Match 13.1%; Score 135; DB 3; Length 332;
Best Local Similarity 27.6%; Pred. No. 0.00084;
Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;

QY 6 HQMISNNKTERSTAPLNTQISALRNDKPLQPPAPANQNSQNTLQTPPIAP 65
DB 7 HMQVDNTRRGRGLAALAIAAAS-ASLVTVAVPATANADP-----BPAPVPPT 57
QY 66 AKP-----AAPRLDRESFGVENKLIPSGSPASSTPLPDGTGNTNNRAVTVSQ 121
DB 58 AASPSTAAAPP-----ATPVAPPPPAANTEN---AQSGDP 93
QY 122 GSNSSADPKAPPPPVSSGEP-PTLGENPDG 152
DB 94 WAAPPADPNAPPPVFIAPNAPQPVRIIDNPVG 125

RESULT 14
US-08-818-111-53
; Sequence 53, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA: US/08/819,111  
 FILING DATE: 13-MAR-1997  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maki, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.417C6  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 53:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 332 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-818-111-53

Query Match 13.1%; Score 135; DB 4; Length 332;  
 Best Local Similarity 27.6%; Pred. No. 0.00084;  
 Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;  
 QY 6 HICNISNNKTERSTAPLNTQISALRNDEPKLPQPPAPANODQSSQNTRLQPTTPIAP 65  
 DB 7 HMQVDPNLTRKGRGLAALAIAAMAS-ASLVTVAVPATANADP-----EPAPPVPTT 57  
 QY 66 AKPP-----AAPRPLDRESFGVENKLIPSVGSPASSTLPPDGTGNGSTPNNRNAVTEVSQ 121  
 DB 58 AASPSTAAAPPAP-----ATPVAPPPFAAANTPN---AQPQDP 93  
 QY 122 GSNSSSADPKAPPPPPVSSGEP-PTLGENPDG 152  
 DB 94 NAAAPPADPNAPPPPPVIAPNAPQVVRIDNPVG 125

RESULT 15  
 US-09-056-556-53  
 Sequence 53, Application US/09056556  
 Patent No 6350456  
 GENERAL INFORMATION:  
 APPLICANT: Reed, Steven G.  
 APPLICANT: Skeiky, Yasir A.W.  
 APPLICANT: Dillon, Davin C.  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND  
 NUMBER OF SEQUENCES: 241  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: SEED and BERRY LLP  
 STREET: 6300 Columbia Center, 701 Fifth Avenue  
 CITY: Seattle  
 STATE: Washington  
 COUNTRY: USA  
 ZIP: 98104-7092  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/056,556  
 FILING DATE: 07-APR-1998  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Maki, David J.  
 REGISTRATION NUMBER: 31,392  
 REFERENCE/DOCKET NUMBER: 210121.457  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (206) 622-4900  
 TELEFAX: (206) 682-6031  
 INFORMATION FOR SEQ ID NO: 53:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 332 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-056-556-53  
 Query Match 13.1%; Score 135; DB 4; Length 332;  
 Best Local Similarity 27.6%; Pred. No. 0.00084;  
 Matches 42; Conservative 15; Mismatches 57; Indels 38; Gaps 6;  
 QY 6 HICNISNNKTERSTAPLNTQISALRNDEPKLPQPPAPANODQSSQNTRLQPTTPIAP 65  
 DB 7 HMQVDPNLTRKGRGLAALAIAAMAS-ASLVTVAVPATANADP-----EPAPPVPTT 57  
 QY 66 AKPP-----AAPRPLDRESFGVENKLIPSVGSPASSTLPPDGTGNGSTPNNRNAVTEVSQ 121  
 DB 58 AASPSTAAAPPAP-----ATPVAPPPFAAANTPN---AQPQDP 93  
 QY 122 GSNSSSADPKAPPPPPVSSGEP-PTLGENPDG 152  
 DB 94 NAAAPPADPNAPPPPPVIAPNAPQVVRIDNPVG 125

Search completed: November 13, 2003, 09:28:22  
 Job time : 29.6848 secs

TREATM

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:27:44 ; Search time 52.0856 Seconds  
(without alignments)  
679.968 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392  
Perfect score: 1028  
Sequence: 1 TIVSFHIQINNNKTERSTA.....LPDEKFTGAGGFGQNP 194

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1028	100.0	1426	15	US-10-322-579-15
2	178	17.3	35	15	Sequence 15, Appl
3	167.5	16.3	707	11	US-09-919-039-278
4	162.5	15.8	574	12	US-10-168-097A-76
5	162.5	15.8	574	12	US-10-239-431A-38
6	158	15.4	802	9	US-09-823-240-2
7	154.5	15.0	731	14	US-10-086-464-17
8	152	14.8	509	12	US-10-021-660-86
9	151.5	14.7	503	14	US-10-078-547-2
10	150.5	14.6	533	11	US-09-922-226-2
11	150.5	14.6	533	15	US-10-038-184-2
12	150.5	14.6	800	12	US-09-929-586-32198
13	150	14.6	5179	9	US-09-923-217-1068
14	150	14.6	5179	10	US-09-833-263-1068
15	150	14.6	5179	14	US-10-025-380-1068

16	149	14.5	507	14	US-10-078-547-24
17	148.5	14.4	731	14	US-10-086-464-8
18	144.5	14.1	351	12	US-10-239-431A-22
19	144.5	14.1	542	12	US-10-239-431A-10
20	144.5	14.1	844	15	US-10-156-761-7663
21	142.5	13.9	721	14	US-10-086-464-5
22	141.5	13.8	416	15	US-10-043-487-282
23	141.5	13.8	505	12	US-10-168-097A-12
24	141.5	13.8	505	12	US-10-239-431A-32
25	141	13.7	666	9	US-09-791-171-70
26	141	13.7	666	12	US-09-804-980-70
27	141	13.7	666	15	US-10-080-170-640
28	139.5	13.6	505	12	US-10-168-097A-56
29	139.5	13.6	505	12	US-10-239-431A-36
30	139	13.5	501	12	US-10-168-097A-46
31	139	13.5	501	12	US-10-239-431A-35
32	139	13.5	638	15	US-10-038-010-4
33	138	13.4	377	15	US-10-149-819-7
34	137.5	13.4	647	14	US-10-086-464-2
35	137.5	13.4	647	14	US-10-086-464-4
36	136.5	13.3	523	12	US-10-017-161-1982
37	136	13.2	633	12	US-10-168-097A-66
38	136	13.2	633	12	US-10-239-431A-37
39	135	13.1	332	12	US-10-084-843-53
40	135	13.1	332	12	US-10-193-002-53
41	135	13.1	332	12	US-10-098-732A-41
42	134.5	13.1	200	9	US-09-925-297-700
43	134.5	13.1	668	10	US-09-925-300-1633
44	134.5	13.1	668	12	US-10-240-154-20
45	134.5	13.1	704	12	US-10-240-154-18

## ALIGNMENTS

RESULT 1  
US-10-322-579-15  
; Sequence 15, Application US/10322579  
; Publication No. US2003014413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMES, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; FILE REFERENCE: 060361  
; CURRENT APPLICATION NUMBER: US/10/322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 15  
; LENGTH: 1426  
; TYPE: PRT  
; ORGANISM: Human lgs/bcl9  
US-10-322-579-15

Query Match 100.0%; Score 1028; DB 15; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 8.7e-59;  
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TIVSFHIQINNNKTERSTAPLNTQISALNDPKLPQPPANQDQNSQNTLQPTP 60
Db	199	TIVSFHIQINNNKTERSTAPLNTQISALNDPKLPQPPANQDQNSQNTLQPTP 258
Qy	61	PIAPAPKAPAPPLDRESGVNKLIVSGVSPASSTPLPPGTCGNSQNTNPAVTVS 120
Db	259	PIAPAPKAPAPPLDRESGVNKLIVSGVSPASSTPLPPGTCGNSQNTNPAVTVS 318



TYPE: PRT  
ORGANISM: Schizosaccharomyces pombe  
US-10-239-431A-38

Query Match 15.8%; Score 162.5; DB 12; Length 574;  
Best Local Similarity 30.9%; Pred. No. 0.0045;  
Matches 50; Conservative 21; Mismatches 64; Indels 27; Gaps 7;  
QY 7 IQNISNNKTERSTA-----PLNTQISA---LRNDKPLFQPPAPANQDQ---NSSQNT 54  
DB 275 IAPVSMNPAINSTSKPPLPPSSVSAALAAANKRPPPPPPSRNRGRKPPGNGSSNS 334  
QY 55 RIQPTFP-----IPAPAPKPAAPPLDRESQGVENKLIPIVSGSPA-SSTPLPPDG 104  
DB 335 SLPPPPPPPSRNAAGSIPLPQGRSAPPPPPPPSAPSTGRQPPPLSSSRVSNPPAPPPA 394  
QY 105 TGFNSTNNRAVTFVSGSNSSADPKAPPPVSSGSEPTL 146  
DB 395 IFRSAP---ALPLGNASRTST--PPVTFPSLPPPSAPPSL 431

RESULT 6  
US-09-823-240-2

Sequence 2, Application US/09823240  
Patent No. US2002048813A1  
GENERAL INFORMATION:  
APPLICANT: Frank B. Gertler  
APPLICANT: James E. Bear  
APPLICANT: Jurgen Wehland  
APPLICANT: Joseph Loureiro  
TITLE OF INVENTION: Motility  
FILE REFERENCE: M0656/7064 (HCL)  
CURRENT APPLICATION NUMBER: US/09/823,240  
PRIORITY FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 60/194,564  
PRIORITY FILING DATE: 2000-04-03  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2  
LENGTH: 802  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-823-240-2

Query Match 15.4%; Score 158; DB 9; Length 802;  
Best Local Similarity 26.3%; Pred. No. 0.012;  
Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;  
QY 13 NKTERSTAPLNT-----QISALNDP----- 33  
DB 348 NKNSRSPSVNTSSQPPAAKSCAWPTNSFPLPPFPMISSPGKATGPRVLPVCVS 407  
QY 34 KPLFQPPAPANQDQ-----SSQNTLRQFTTPIAPAPKPAAPRPL----- 76  
DB 408 SEVQMPSPSPAPNGSLDSVTYVSPPTSGFAAPPPPPPPPPPPPPPPPLPLAS 467  
QY 77 -----DRESQGVENKLI-----PSVGSFSS--TLPDP--GTGNGTNNRAV-- 116  
DB 468 LSHCGSQASPPRGTPLASTPSSKPSVLPSPSAGAPASAEPLNPELDSSASEPGLQAS 527  
QY 117 ----TPVSGSNSSADPKAPPPVSSGSE-----PPTLGENP 150  
DB 528 QPAESPTQD--LVLGPPAPPPPPPLPSGPAYASALPPPPPPPP 569

RESULT 7  
US-10-086-464-17

Sequence 17, Application US/10086464  
Publication No. US20020199218A1  
GENERAL INFORMATION:  
APPLICANT: GORING, DAPHNE R. et al.  
TITLE OF INVENTION: PROLINE-RICH EXTENSIN-LIKE RECEPTOR KINASES

FILE REFERENCE: P 25,762-A USA  
CURRENT APPLICATION NUMBER: US/10/086,464  
CURRENT FILING DATE: 2002-02-28  
PRIOR APPLICATION NUMBER: US 10/069,304  
PRIOR FILING DATE: 2002-02-19  
PRIOR APPLICATION NUMBER: PCT/CA00/00966  
PRIOR FILING DATE: 2000-08-18  
PRIOR APPLICATION NUMBER: US 60/149,466  
PRIOR FILING DATE: 1999-08-19  
PRIOR APPLICATION NUMBER: US 60/159,122  
PRIOR FILING DATE: 1999-10-13  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 17  
LENGTH: 731  
TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-10-086-464-17

Query Match 15.0%; Score 154.5; DB 14; Length 731;  
Best Local Similarity 27.6%; Pred. No. 0.019;  
Matches 47; Conservative 20; Mismatches 68; Indels 35; Gaps 7;  
QY 15 TERSTAPLNTQISALRNDKPLFQPPAPANQDQSSQNTLRQFTTPIAPAPK--PAAP 72  
DB 29 TSSPPAPPLSPPLPPLSSPPPLSAPTASPPPLSVSPSPPIESPPPLLESPP 88  
QY 73 PRPLDRESQGVENKLIPIVSGSPASSTPL-----PPDGTGPNST--PNNRAVTFVSG-- 122  
DB 89 PPPLSPSPPE-----SPHVSAPSGSPPLPPLPAKPSPPSPSPSETVPPGNTISPPPSL 143  
QY 123 -----SNSSADPKAPP-----PPVSSGSEPTLGENPGLSQ 155  
DB 144 PSESTPPVNTASPPSPSPRRSRGKPSFPPTINS--SPNTSPNTPLSP 192

RESULT 8

US-10-021-660-86  
Sequence 86, Application US/10021660  
Publication No. US20030152926A1  
GENERAL INFORMATION:  
APPLICANT: Murray, Richard  
APPLICANT: Glynn, Richard  
APPLICANT: Watson, Susan R.  
APPLICANT: EOS Biotechnology, Inc.  
TITLE OF INVENTION: No. US20030152926A1el Methods of Diagnosis of Angiogenesis,  
TITLE OF INVENTION: Compositions and Methods of Screening for Angiogenesis  
TITLE OF INVENTION: Modulators  
FILE REFERENCE: 018501-000710US  
CURRENT APPLICATION NUMBER: US/10/021,660  
CURRENT FILING DATE: 2001-12-06  
PRIOR APPLICATION NUMBER: US/09/784,356  
PRIOR FILING DATE: 2001-02-14  
PRIOR APPLICATION NUMBER: US 09/637,977  
PRIOR FILING DATE: 2000-08-11  
NUMBER OF SEQ ID NOS: 135  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 86  
LENGTH: 509  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-021-660-86

Query Match 14.8%; Score 152; DB 12; Length 509;  
Best Local Similarity 27.0%; Pred. No. 0.019;  
Matches 60; Conservative 13; Mismatches 65; Indels 84; Gaps 10;

QY 28 ALRNDKPLP-----QQPPAPANQDQSS-----SQNTLRQFTTPIAPAPKPA 70  
DB 93 ALHDAFRAVPGLLALGLVTGQADQRFAGARQQQQPQQRQVPAACQPPVPRHQVHP 152  
QY 71 APPRLDRESQGVENKLIPIVSG-----SPASSTPL--PPDGTGNS----- 109

Db 153 APPPPPPSRAGSAGALPCAGHTRRRRTTSRRSSPPLSGPPGRASPRGARPPPLLRRAA 212  
 QY 110 -TNNRVAVTVSGNSSSADPKAPPPPVSSGE-----PP-----TL 146  
 Db 213 FTSPRALAP-----AAASPPPPPPPPGREGKRRKPPGSGSTQTSGAAVAAL 265  
 QY 147 GENPDGLSQEQLHREHRSLOTLRIQMLFPDEKFTGAQSG 188  
 Db 266 GSSPG-----RRRLPLLLRVGR-----PRSGAASG 291

RESULT 9  
 US-10-078-547-2  
 ; Sequence 2, Application US/10078547  
 ; Publication No. US2002019211A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Narayanasamy Ramesh  
 ; APPLICANT: Miguel A. de la Fuente  
 ; APPLICANT: Ines M. Anton  
 ; APPLICANT: Ralf S. Geha  
 ; TITLE OF INVENTION: WIP, A WASP-Associated Protein  
 ; FILE REFERENCE: 1242.1022-005  
 ; CURRENT APPLICATION NUMBER: US/10/078, 547  
 ; PRIOR FILING DATE: 2002-02-19  
 ; PRIOR APPLICATION NUMBER: 09/599,287  
 ; PRIOR FILING DATE: 2000-06-22  
 ; PRIOR APPLICATION NUMBER: PCT/US98/27501  
 ; PRIOR FILING DATE: 1998-12-22  
 ; PRIOR APPLICATION NUMBER: 60/101,457  
 ; PRIOR FILING DATE: 1998-09-23  
 ; PRIOR APPLICATION NUMBER: 60/068,533  
 ; PRIOR FILING DATE: 1997-12-23  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 503  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 US-10-078-547-2

Query Match 14.7%; Score 151.5; DB 14; Length 503;  
 Best Local Similarity 26.5%; Pred. No. 0.021; Indels 57; Gaps 11;  
 Matches 60; Conservative 25; Mismatches 84  
 QY 12 NKK-----TRSTAPLNTQISALRNDKPLFQQPPAPANQDONS-----SQNTRL-QPT 59  
 Db 289 NKKPVVSTPRGSAHRPH---LRPPPSRGPPPLPPSSSGNDTPRLPQNLSSST 345  
 QY 60 PPTAPAPKAPAPRLPDRSPGVENKLIPIVSGPASPSTPLPPDGTGPNSTNNRAV---116  
 Db 346 PPLSPGRSGPLPPPSRPPPPVRD-----PGRSGPLPPPPVSRNGSTSRALPAT 398  
 QY 117 --TPVSGNSSSADPK-----APPPPVSGSPPTLGENPDGLSQEQLHRE-- 161  
 Db 399 POLPSRSGVDSFRSGFRPLPPDRPSAGAPPPTSTSRNGFQDSP---CEDENESREY 455  
 QY 162 -----ERSLOTLRIQMLFPDEKFTGA---QSGGPOQNP 194  
 Db 456 FHPISDLPPPEPVYVQTKSYPSKLANNESR-SGSNRREGGFPPLPP 500

RESULT 10  
 US-09-922-226-2  
 ; Sequence 2, Application US/09922226  
 ; Publication No. US20030077664A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhao, Yi  
 ; APPLICANT: Thacher, Scott M.  
 ; APPLICANT: Xiao, Jia-Hao  
 ; APPLICANT: Kusari, Jyotirmoy  
 ; APPLICANT: Chandraratna, Roshantha A.  
 ; TITLE OF INVENTION: Methods of Screening For Compounds That  
 ; Modulate Hormone Receptor Activity

FILE REFERENCE: P-AR 4681  
 CURRENT APPLICATION NUMBER: US/09/922,226  
 CURRENT FILING DATE: 2002-01-09  
 PRIOR APPLICATION NUMBER: US 60/284,797  
 PRIOR FILING DATE: 2001-04-18  
 NUMBER OF SEQ ID NOS: 191  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 533  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-922-226-2

Query Match 14.6%; Score 150.5; DB 11; Length 533;  
 Best Local Similarity 31.0%; Pred. No. 0.025; Indels 33; Gaps 5;  
 Matches 40; Conservative 11; Mismatches 45  
 QY 31 NDPKPLPQ--QPPAPANQDONSQNTLRQTPFPAPAP-----PKPAAPPRPLDRESP 81  
 Db 86 SSNPPLFQGVPPSPFG-----PPLPSTAPSLGSGGAPPMPMPPLGSPFP 134  
 QY 82 GVENKLIPIVSGPASPSTPLPPDGTGPNSTNNRAVTPVSGNSSSSADPKAP-----133  
 Db 135 VISS---SMGSGPLPPAPPGFSGVPSSPQINSTVSLPGSGGPPEDVKPVLGVRGLH 190  
 QY 134 -PPPVSSG 141  
 Db 191 CFPFPGPG 199

RESULT 11  
 US-10-098-184-2  
 ; Sequence 2, Application US/10098184  
 ; Publication No. US2003010533A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pfahl, Magnus  
 ; APPLICANT: Tachdjian, Catherine  
 ; APPLICANT: Al-Shamma, Hussien A.  
 ; APPLICANT: Fanjul, Andrea  
 ; APPLICANT: Pleyner, David P.M.  
 ; APPLICANT: Spruce, Lyle W.  
 ; APPLICANT: Fine, Richard  
 ; APPLICANT: Zapf, James W.  
 ; TITLE OF INVENTION: RXR ACTIVATING MOLECULES  
 ; FILE REFERENCE: 13099.0016U2  
 ; CURRENT APPLICATION NUMBER: US/10/098,184  
 ; CURRENT FILING DATE: 2002-03-08  
 ; PRIOR APPLICATION NUMBER: 60/274,342  
 ; PRIOR FILING DATE: 2001-03-08  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 533  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence; No. US2003010533A1e =  
 ; OTHER INFORMATION: synthetic construct  
 US-10-098-184-2

Query Match 14.6%; Score 150.5; DB 15; Length 533;  
 Best Local Similarity 31.0%; Pred. No. 0.025;  
 Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;  
 QY 31 NDPKPLPQ--QPPAPANQDONSQNTLRQTPFPAPAP-----PKPAAPPRPLDRESP 81  
 Db 86 SSNPPLFQGVPPSPFG-----PPLPSTAPSLGSGGAPPMPMPPLGSPFP 134  
 QY 82 GVENKLIPIVSGPASPSTPLPPDGTGPNSTNNRAVTPVSGNSSSSADPKAP-----133  
 Db 135 VISS---SMGSGPLPPAPPGFSGVPSSPQINSTVSLPGSGGPPEDVKPVLGVRGLH 190  
 QY 134 -PPPVSSG 141



```
Db      191 CPTTGGG 199
      |||||
RESULT 12
US-10-029-386-32198
; Sequence 32198, Application US/10029386
; Publication No. US200301947041
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029, 386
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32198
; LENGTH: 800
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007663.28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
US-10-029-386-32198
Query Match      14.6%; Score 150.5; DB 12; Length 800;
Best Local Similarity 33.6%; Pred. No. 0.038;
Matches 41; Conservative 10; Mismatches 52; Indels 19; Gaps 5;
QY      33 PRLPQQPAPANDQNSQNTRLQPTTP-----IPAPAKPAAPRPLDRESGVENKLI 88
Db      609 PRLPQLPPLPPSPS-----LPPPPSPSLPPLPPPPSLPPLPPSPSLPSP 660
QY      89 PSVGSPPASTLPPDGTGNSTPNNAVTPVSGQNSSSADPKAPPVPVSSGEPPTLGE 148
Db      861 PSLSP--SLPLP--PPSLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLPPLP 713
QY      149 NP 150
Db      714 SP 715
RESULT 13
US-09-922-217-1068
; Sequence 1068, Application US/09922217
; Patent No. US200200764141
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-217-1068
Query Match      14.6%; Score 150; DB 10; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.27;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;
QY      15 TERSTAPLNTQISALNDKPLQCPQAPANQDQNSQNTLQPTTPAPAKPAAP-- 72
Db      1428 TTTTPPTTSPPIITTTTLPITTTTSPPISTTTTTPPT-TTPSPPTTSPPTTSP 1486
QY      73 -----PRLDRESGVENKLI SVGSPPASTLPPDGT-----GNSTPNNR 114
Db      1487 PTTTTPPTTTPPTTSPPTTTP-----PASTTLPPTTTPSPPTTTPPTTSP 1542
QY      115 AVTPVSGQNSSSADPKAPP-----PPVSSGEPPT 145
Db      1543 TTTTTPPTTSTTTLPTTTPPTTTPPTTTPPTTTPPTTTPPTTSPPT 1582
RESULT 14
US-09-833-263-1068
; Sequence 1068, Application US/09833263
; Patent No. US20020110547A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Meagher, Madeleine J.
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND
; FILE REFERENCE: 210121.471C12
; CURRENT APPLICATION NUMBER: US/09/833.263
; CURRENT FILING DATE: 2001-04-10
; NUMBER OF SEQ ID NOS: 1093
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-263-1068
Query Match      14.6%; Score 150; DB 10; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.27;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;
QY      15 TERSTAPLNTQISALNDKPLQCPQAPANQDQNSQNTLQPTTPAPAKPAAP-- 72
Db      1428 TTTTPPTTSPPIITTTTLPITTTTSPPISTTTTTPPT-TTPSPPTTSPPTTSP 1486
QY      73 -----PRLDRESGVENKLI SVGSPPASTLPPDGT-----GNSTPNNR 114
Db      1487 PTTTTPPTTTPPTTSPPTTTP-----PASTTLPPTTTPSPPTTTPPTTSP 1542
QY      115 AVTPVSGQNSSSADPKAPP-----PPVSSGEPPT 145
Db      1543 TTTTTPPTTSTTTLPTTTPPTTTPPTTTPPTTTPPTTTPPTTSPPT 1582
RESULT 15
US-10-025-380-1068
; Sequence 1068, Application US/10025380
; Publication No. US20020182191A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole Lynn
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.471C13
; CURRENT APPLICATION NUMBER: US/09/922.217
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 1124
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068
Query Match      14.6%; Score 150; DB 10; Length 5179;
Best Local Similarity 25.6%; Pred. No. 0.27;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;
QY      15 TERSTAPLNTQISALNDKPLQCPQAPANQDQNSQNTLQPTTPAPAKPAAP-- 72
Db      1428 TTTTPPTTSPPIITTTTLPITTTTSPPISTTTTTPPT-TTPSPPTTSPPTTSP 1486
QY      73 -----PRLDRESGVENKLI SVGSPPASTLPPDGT-----GNSTPNNR 114
Db      1487 PTTTTPPTTTPPTTSPPTTTP-----PASTTLPPTTTPSPPTTTPPTTSP 1542
QY      115 AVTPVSGQNSSSADPKAPP-----PPVSSGEPPT 145
Db      1543 TTTTTPPTTSTTTLPTTTPPTTTPPTTTPPTTTPPTTTPPTTSPPT 1582
```

```
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yugu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: OF COLON CANCER AND METHODS FOR THEIR USE
; CURRENT APPLICATION NUMBER: US/10/025,380
; CURRENT FILING DATE: 2001-12-19
; NUMBER OF SEQ ID NOS: 1129
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1068
; LENGTH: 5179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-025-380-1068

Query Match      14.6%; Score 150; DB 14; Length 5179;
Best Local Similarity 25.6%; Pred.No. 0.27;
Matches 41; Conservative 20; Mismatches 65; Indels 34; Gaps 5;

QY      15 TERSTAPLNTQIGALRNDEKELPOQPPAPANODQNSQNTRLQPTPIPAKPAAP-- 72
Db      1428 TTTTPPTTTPSPPIITTTTTLPTTTPSPPISTTTTPPPT-TTPSPFTTTPSPFTTTEP 1486

QY      73 -----PRLDRESEGVENKLIPIVSGSPASSTLPPDGT-----GENSTPNR 114
Db      1487 PTTTTTPPTTTPSPMTTTP-----PASTTTLPTTTPSPPTTTTTPPTTTPGPP 1542

QY      115 AVTPVSGSNSSGADPKAPP-----PPPVSSGEPPT 145
Db      1543 TTTTTPPTSTTLPTTTPGPPPTTTTTPPTTTPGPPPT 1582
```

Search completed: November 13, 2003, 09:39:45  
Job time : 53.0856 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 26.4202 Seconds  
(without alignments)  
706.153 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028  
Sequence: 1 TIVSFH1QINSNKTERSTA.....LFPDEKFTGAQSGPQNP 194

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	177.5	17.3	473	2 S50755	hypothetical prote
2	168.5	16.4	876	2 T49801	hypothetical prote
3	167.5	16.3	707	2 A46302	PTB-associated spl
4	162.5	15.8	574	2 T13556	Wiskott-Aldrich sy
5	162.5	15.8	574	2 T38819	wiskott-aldrich sy
6	162.5	15.8	1151	2 T18835	high molecular mas
7	161.5	15.7	815	2 B56708	extracellular sign
8	160	15.6	839	2 F75518	hypothetical prote
9	158.5	15.4	351	2 S50754	hypothetical prote
10	158.5	15.4	1188	2 S49915	extensin-like prot
11	157.5	15.3	464	2 S22697	extensin - Volvox
12	156	15.2	449	2 S16748	proline-rich prote
13	154.5	15.0	731	2 T04455	hypothetical prote
14	152.5	14.8	379	2 T05441	proline-rich prote
15	152.5	14.8	379	2 D85257	extensin-like prote
16	150.5	14.6	251	1 PTHUPF	salivary proline-r
17	150.5	14.6	533	2 S37781	retinoid X recepto
18	150	14.6	3020	2 A43932	mucin 2 precursor,
19	149	14.5	198	2 D39149	proline-rich prote
20	149	14.5	302	2 T32111	hypothetical prote
21	149	14.5	436	2 J01696	pistil extensin-li
22	148.5	14.4	403	2 G86441	unknown protein [i
23	148.5	14.4	403	2 S2796	prpL2 protein - hu
24	148.5	14.4	731	2 B86369	hypothetical prote
25	147	14.3	348	1 OZZQBK	circumsporozoite p
26	147	14.3	416	2 T34279	hypothetical prote
27	146.5	14.3	393	2 P00479	pistil extensin-li
28	146.5	14.3	708	2 D56711	hypothetical prote
29	146	14.2	240	2 B24964	proline-rich prote

#### ALIGNMENTS

##### RESULT 1

S50755  
hypothetical protein VSP-3 - Chlamydomonas reinhardtii  
C:Species: Chlamydomonas reinhardtii  
C>Date: 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-2000  
C:Accession: S50755  
R:Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Har  
Plant Mol. Biol. 26, 947-960, 1994  
A:Title: Domain conservation in several volvoclean cell wall proteins.  
A:Reference number: S50754; NUID:95093034; PMID:8000007  
A:Accession: S50755  
A:Molecule type: mRNA  
A:Status: preliminary  
A:Residues: 1-473 <WQE>  
A/Cross-references: EMBL:L29029; NID:G530875; PID:NAB53953.1; PID:G530876

Query Match 17.3%; Score 177.5; DB 2; Length 473;  
Best Local Similarity 30.5%; Pred. No. 0.0019;  
Matches 46; Conservative 19; Mismatches 59; Indels 27; Gaps 5;

QY 14 KTERSTAPLNTQISALRNDKPLPQQPPAPANQDNSSQNTRLQPTPIIPAPKPAAPP 73  
DB 330 KASPSPSPSVQPAKSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 376  
QY 74 RPLDRESFGVENKLLPSVGSFASSTPLPDGTGPNSTPN-NRAVTPVSGSNSSGADPKA 132  
DB 377 SP---SPSPSPKPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSKA 432  
QY 133 -----PPPPVSSSGEPPTLGENPDGLSQ 155  
DB 433 SPSPAKKPSPPPPVEGAPPPI-EGPPPMEE 462

##### RESULT 2

T49801  
hypothetical protein B11B22.30 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C>Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 02-Jun-2000  
C:Accession: T49801  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakat  
submitted to the Protein Sequence Database, May 2000  
A:Reference number: 225022  
A:Accession: T49801  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-876 <SCH>  
A/Cross-references: EMBL:AL356834; GSPDB:GN00116; NCSP:B11B22.30  
A:Experimental source: BAC clone B11B22; strain OR74A  
A:Gene: NCSP:B11B22.30  
A:Map position: 6  
A:Introns: 75/3; 190/1; 449/3

30 146 14.2 439 2 S51939  
31 146 14.2 476 2 T27051  
32 146 14.2 1137 2 A86335  
33 145.5 14.2 260 2 S22373  
34 145 14.1 534 2 S21961  
35 144.5 14.1 542 2 A44358  
36 144.5 14.1 577 2 T09024  
37 144.5 14.1 760 2 F86387  
38 144.5 14.1 1611 2 T18236  
39 144 14.0 535 1 S76953  
40 144 14.0 817 2 S51342  
41 143.5 14.0 446 2 S34418  
42 143.5 14.0 488 2 D41727  
43 143.5 14.0 520 2 I84718  
44 143 13.9 300 2 S19560  
45 143 13.9 332 1 OZZQBK

chitinase (EC 3.2.  
hypothetical prote  
T20H2.9 protein -  
proline-rich prote  
proline-rich prote  
zyxin - chicken  
proline-rich prote  
probable Pto kinase  
hypothetical prote  
protein kinase [EC  
VapA protein, yeast  
H-2 region II bind  
retinoid X recepto  
RXR-beta1 isoform  
proline-rich prote  
circumsporozoite p



## RESULT 6

T18535  
high molecular mass nuclear antigen - chicken (fragment)  
C:Species: Gallus gallus (Chicken)  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18535  
R:Shimada, K.; Harata, M.; Mizuno, S.  
J. Cell Sci. 110, 3031-3041, 1997  
A:Title: A nuclear matrix-associated high molecular mass nuclear antigen, HMNA, of chick  
A:Reference number: 218955; MUID:9803440; PMID:9365273  
A:Accession: T18535  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1151 <SHI>  
A:Cross-references: EMBL:D88440; NID:d1177138; PID:d1025045; PIDN:BAA24137.1

Query Match 15.8%; Score 162.5; DB 2; Length 1151;  
Best Local Similarity 33.1%; Pred. No. 0.033;  
Matches 45; Conservative 12; Mismatches 40; Indels 39; Gaps 6;

QY 33 PKPLPQQPPAPANQDQSSQNTLQPT-PIPAAPAPAPAPPP-----LDRESFGVE 84  
DB 3 PTPPTPRNDPTP-----PPAPSPAPAPAPAPTAPPRPKWPIAELHPAAPP 49  
QY 85 NKLIPIVSGSPASSTPLPPDGTGPNSTNNRAVTPVSGSNSSADPKAP---PPPPVSSG 141  
DB 50 PKWPIIGAP-----PPPGTER-----TPESKPTDGDAAPKASAEITSPPPASP 94  
QY 142 EPTLGENPDGLSQEQ 157  
DB 95 SPDPGPKAPSGAGEAE 110

## RESULT 7

B56708  
extracellular signal-regulated kinase 5 - human  
C:Species: Homo sapiens (man)  
C>Date: 19-Oct-1995 #sequence\_revision 19-Oct-1995 #text\_change 24-Sep-1999  
C:Accession: B56708  
R:Zhao, G.; Bao, Z.Q.; Dixon, J.E.  
J. Biol. Chem. 270, 12665-12669, 1995  
A:Title: Components of a new human protein kinase signal transduction pathway.  
A:Reference number: A56708; MUID:195279403; PMID:7759517  
A:Accession: B56708  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-815 <ZHO>  
A:Cross-references: GB:U25278; NID:9837260; PIDN:AA81391.1; PID:9837261  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog  
C:Keywords: ATP  
F:52-315/Domain: protein kinase homology <KIN>  
F:60-68/Region: protein kinase ATP-binding motif

Query Match 15.7%; Score 161.5; DB 2; Length 815;  
Best Local Similarity 32.1%; Pred. No. 0.027;  
Matches 54; Conservative 7; Mismatches 64; Indels 43; Gaps 7;

QY 17 RSTAPLNTQ1SALRNDPKLP-----QQPPAPANQDQSSQNTLQPTTPIP 63  
DB 576 RPAAPALTSVPAPAPAPTPTPTVQPTSPPPGLAQTGPQSGAGSTGSPVQACPPR 635  
QY 64 APAPKAPAPPPFLDRESFG-----VENKLIPSVGSPASSTP-----LPPDGT 105  
DB 636 GPAPHPTGPGPIVPVAPPQIATSTSLAAQSLVPPFGLGSGTGVLPYFPFGLPPDA 695  
QY 106 G--PNS---PPNNRAVTPVSGSNSSADPKAPPPPPVSSGPPTLG 147  
DB 696 GGAPQSSMSEPDNLVT---QQLSKQVEDPL-----PPVSGTPKSG 737

## RESULT 8

F75518  
hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 31-Mar-2000  
C:Accession: F75518  
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: F75518  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-839 <WHI>  
A:Cross-references: GB:AB0001904; GB:AB000513; NID:96458129; PIDN:AAE10038.1; PID:96458129  
A:Experimental source: strain R1  
C:Genetics:

A:Gene: DR0458  
C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology

Query Match 15.6%; Score 160; DB 2; Length 839;  
Best Local Similarity 31.3%; Pred. No. 0.033;  
Matches 52; Conservative 13; Mismatches 79; Indels 22; Gaps 7;

QY 33 PKP-LPQQPPAPANQDQSSQNTLQPT---TPPIPAPA---PKPAAP-PRFLDRESFGVE 84  
DB 222 PTFNAPAQTPAPATQAPAAQTPTAQAPATQTPTATPAAQRPAGGAPSPAPAPANAPA 281  
QY 85 NKLIPIVSGSPASSTPLPPDGTGPNSTNNRAVTPVSGS-NSSSADPKAPPPPPVSSGSP 143  
DB 282 GSVVPEATVPESSTPAFSAQTPTPTTRETATQTEASPAFNSAAAFAPASEPV-AGRP 340  
QY 144 PTLGENPDGLSQEQLEHREHSLOTLRDIQRMFLPDEKEFTGAOSGG 189  
DB 341 GTAASSPESASPVTVTPTGET-----PDTASAGTFSAG 374

## RESULT 9

S50754  
hypothetical protein WP6 - Chlamydomonas eugametos  
C:Species: Chlamydomonas eugametos  
C>Date: 14-Jul-1995 #sequence\_revision 21-Jul-1995 #text\_change 21-Jul-2000  
C:Accession: S50754  
R:Woesener, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Hart  
Plant Mol. Biol. 26, 947-960, 1994  
A:Title: Domain conservation in several volvocalean cell wall proteins.  
A:Reference number: S50754; MUID:95093034; PMID:8000007  
A:Accession: S50754  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-351 <MOE>  
A:Cross-references: EMBL:L29028; NID:9530877; PIDN:AAB53954.1; PID:9530878

Query Match 15.4%; Score 158.5; DB 2; Length 351;  
Best Local Similarity 30.9%; Pred. No. 0.017;  
Matches 46; Conservative 15; Mismatches 71; Indels 17; Gaps 5;

QY 9 MISNNKTERSTAPLNTQ-----ISALRNDPKLPQQPPAPANQDQSS-----SONTRLOPT 59  
DB 144 SVYANCIDTRPAPNCSTFNVTSTVITPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 203  
QY 60 PIPAPAPKPAAPPRFLDRESFGVENKLIPSVGSPASSTPLPPDGTGPNSTNNRAVTP 118  
DB 204 PSPKASPSPKASPSAPSP 258  
QY 119 VSGSNSSADPKAPP---PPPVSGEPT 145  
DB 259 SPKASPPSPASPSAPSP 287

## RESULT 10



```

A:Note: F7K2.50
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match      14.8%; Score 152.5; DB 2; Length 379;
Best Local Similarity 31.8%; Pred. No. 0.041;
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY 33 PKPLFQQPPAPANQDSSQNTRLQPTPEI---PAPAPKPAAPRPPLDRESGVENKLIIP 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 PQPDPQPPPTPT-----FQAPPANDQPPPPQSTSPF-FVATTPPALPKPLP 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 90 SVGSPASSTPLPDGCTGNSTNNRAVTFVSGNSSSADPKAPPPPPVSSGEPPTLGEN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 PPLSPFQTTTPPPPAITPPPPF---AITE-----PLSPPPPAIT--PPPLATT 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 150 PDGLSQEQ 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 PPALPPKPL 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
D85257
extensin-like protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: D85257
R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
  Harbor, NY, 1999
C:Title: 402, 769-777, 1399
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; PMID:20083488; PMID:10617198
A:Accession: D85257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-379 <STO>
A:Cross-references: GB:NC_001269; NID:97269093; PIDN:CAE79202.1; GSPDB:GN00140
A:Gene: AT4g22470
A:Map position: 4
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match      14.8%; Score 152.5; DB 2; Length 379;
Best Local Similarity 31.8%; Pred. No. 0.041;
Matches 41; Conservative 12; Mismatches 45; Indels 31; Gaps 6;

QY 33 PKPLFQQPPAPANQDSSQNTRLQPTPEI---PAPAPKPAAPRPPLDRESGVENKLIIP 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 51 PQPDPQPPPTPT-----FQAPPANDQPPPPQSTSPF-FVATTPPALPKPLP 98
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 90 SVGSPASSTPLPDGCTGNSTNNRAVTFVSGNSSSADPKAPPPPPVSSGEPPTLGEN 149
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 99 PPLSPFQTTTPPPPAITPPPPF---AITE-----PLSPPPPAIT--PPPLATT 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 150 PDGLSQEQ 158
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 PPALPPKPL 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: November 13, 2003, 09:29:11
Job time : 28.4202 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 : Search time 17.3619 Seconds  
(without alignments)  
525.472 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028  
Sequence: 1 TIVSFHQISNNKTERST.....LFPDEKFTGSGGPGQNP 194

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues  
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1024	99.6	1426	BCL9_HUMAN	Q00512 homo sapien
2	167.5	16.3	707	SPQ_HUMAN	P23246 homo sapien
3	161.5	15.7	815	MK07_HUMAN	Q13164 homo sapien
4	159.5	15.5	555	GPI_CHLSE	Q07566 chlamydomon
5	158.5	15.4	802	ENAH_MOUSE	Q03173 mus musculus
6	156	15.2	449	APC_BRANA	P40603 braasica na
7	151.5	14.7	503	WAF1_HUMAN	Q43516 homo sapien
8	150.5	14.6	253	PRP2_HUMAN	P02812 homo sapien
9	150.5	14.6	533	RXSB_HUMAN	P28702 homo sapien
10	150	14.5	5179	MUC2_HUMAN	Q02817 homo sapien
11	149	14.5	426	EXLP_TOBAC	Q03211 nicotiana t
12	147	14.3	347	CST_FABA	P23033 pisummedium
13	146	14.2	534	APC_ARATH	P40602 arabidopsis
14	146	14.2	2167	SHK1_RAT	Q99448 rattus norv
15	144.5	14.1	339	CSP_FLASE	P08915 plasmodium
16	144.5	14.1	542	ZYX_CHICK	Q04584 gallus gall
17	144	14.0	535	SPKC_SYNY3	P74745 synchocyst
18	144	14.0	817	VRL1_YEAST	P37370 saccharomyc
19	143.5	14.0	520	RXRB_MOUSE	P28704 mus musculus
20	143	13.9	296	PRP3_MOUSE	Q05143 mus musculus
21	143	13.9	1206	FM14_MOUSE	Q05859 mus musculus
22	143	13.9	1468	FMN1_MOUSE	Q05860 mus musculus
23	142.5	13.9	276	PRP1_HUMAN	P10162 homo sapien
24	141.5	13.8	283	EXTN_SORBI	P24152 sorghum bic
25	141.5	13.8	505	WASL_HUMAN	Q00401 homo sapien
26	141.5	13.8	620	EXTN_TOBAC	P13983 nicotiana t
27	140.5	13.7	582	MNT_HUMAN	Q99583 homo sapien
28	140	13.6	806	MK07_MOUSE	Q99583 mus musculus
29	139.5	13.6	268	NO20_MEOTR	P93329 medicago tr
30	139.5	13.6	505	WASL_BOVIN	Q95107 bos taurus
31	139.5	13.6	1664	SLP1_CLOTH	Q06852 clostridium
32	139	13.5	501	WASL_RAT	Q08816 rattus norv
33	138.5	13.5	2161	SHK1_HUMAN	Q99566 homo sapien

## RESULT 1

ID	BCL9_HUMAN	STANDARD	PRT: 1426 AA.
AC	C00512;		
DT	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	15-SEP-2003 (Rel. 42, Last annotation update)		
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).		
GN	BCL9.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Fetal brain;		
RX	MEDLINE=98158621; PubMed=9490669;		
RA	Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,		
RA	Jadavel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,		
RA	Dyer M.J.S.;		
RT	"Molecular cloning of translocation t(1;14) (q21;q32) defines a novel		
RT	gene (BCL9) at chromosome 1q21.";		
RL	Blood 91:1873-1881(1998).		
RN	[2]		
RP	FUNCTION.		
RX	MEDLINE=21952430; PubMed=11955446;		
RA	Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,		
RA	Murphy M., Zuelig S., Basler K.,		
RT	"Wnt/PCP signaling requires Bcl9/legless-mediated recruitment of		
RT	pygopus to the nuclear beta-catenin-TCP complex.";		
CC	Cell 109:47-60(2002).		
CC	-1- FUNCTION: Involved in signal transduction through the wnt pathway.		
CC	-1- SUBUNIT: Binds to beta-catenin (CTNBB1).		
CC	-1- SUBCELLULAR LOCATION: Nuclear (Probable); PYGO1 and PYGO2.		
CC	-1- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,		
CC	testis, ovary and small intestine, and at lower levels in spleen,		
CC	colon and blood.		
CC	-1- DISEASE: Involved in a t(1;14) (q21;q32) chromosomal translocation		
CC	found in a patient with precursor B-cell acute lymphoblastic		
CC	leukemia (ALL). This translocation leaves the coding region		
CC	intact, but may have pathogenic effects due to alterations in the		
CC	expression level of BCL9. Several cases of translocations within		
CC	the 3' untranslated region of BCL9 have been found in B-cell		
CC	malignancies.		
CC	-1- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.		
CC	-1- CAUTION: Ref.1 sequence differs from that shown due to a		
CC	frameshift in position 1391.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).		
CC	-----		

P10220 herpes simp  
P10161 homo sapien  
P05142 mus musculus  
O10341 orgyia pseu  
P03330 simian sarc  
O83384 treponema p  
P49743 rattus norv  
O60610 homo sapien  
O06262 caenorhabdi  
P48634 homo sapien  
Q12446 saccharomyc  
O89935 mus musculus

## ALIGNMENTS



```

DR EMBL; Y13620; CAA73942.1; ALT_FRAME.
DR Genew; HGNC:1008; BCL9.
DR GO; GO:0007048; P:oncogenesis; TAS.
DR GO; GO:0007049; Chromosomal translocation; Proto-oncogene;
KW Nuclear protein; Chromosomal translocation; Proto-oncogene;
KW Wnt signaling pathway.
FT DOMAIN 231 1378 PRO-RICH.
FT DOMAIN 347 377 CTNMB1-BINDING.
FT DOMAIN 331 335 POLY-PRO 1.
FT DOMAIN 514 517 POLY-PRO 2.
FT DOMAIN 900 903 POLY-ALA.
FT DOMAIN 970 973 POLY-PRO 3.
SQ SEQUENCE 1426 AA; 149314 MW; A240A487716B7F1B CRC64;

Query Match 99.6%; Score 1024; DB 1; Length 1426;
Best Local Similarity 99.5%; Pred. No. 4.3e-45;
Matches 193; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TVSFHQNINNKTERSTAPLNTQISALRNDPKLPQPPANQDQSSQNTRELQPTP 60
Db 199 TVSFHQNINNKTERSTAPLNTQISALRNDPKLPQPPANQDQSSQNTRELQPTP 258
QY 61 PIPAPAPKAPRPLDRSFGVENKLIPIESGSPASSTPLPDGTCGNSTNNRAVTVS 120
Db 259 PIPAPAPKAPRPLDRSFGVENKLIPIESGSPASSTPLPDGTCGNSTNNRAVTVS 318
QY 121 QGSNSSADKAPPPVPVSSGEPPTLGENPDGLSQEQLHRSLSQTLRDIQMLFPDEK 180
Db 319 QGSNSSADKAPPPVPVSSGEPPTLGENPDGLSQEQLHRSLSQTLRDIQMLFPDEK 378
QY 181 EFTGAQSGGQQNP 194
Db 379 EFTGAQSGGQQNP 392

RESULT 2
SPFQ HUMAN STANDARD; PRT; 707 AA.
AC P23246; P30808;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Splicing factor, proline- and glutamine-rich (Polypyrimidine tract-
DE binding protein-associated splicing factor) (PTB-associated splicing
DE factor) (PSF) (DNA-binding P52/P100 complex, 100 kDa subunit).
DE SPFQ OR PSF.
GN Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Fetal brain;
RX MEDLINE=93194059; PubMed=8449401;
RA Patton J.G., Porro E.B., Galceran J., Tempst P., Nadal-Ginard B.;
RT "Cloning and characterization of PSF, a novel pre-mRNA splicing
RT factor.";
RL Genes Dev. 7:393-406(1993).
[2]
RN SEQUENCE OF 312-707 FROM N.A.
RC TISSUE=Retal skeletal muscle;
RX MEDLINE=90091812; PubMed=2480877;
RA Gower H.J., Moore S.E., Dickson G., Elsom V.L., Nayak R., Walsh F.S.;
RT "Cloning and characterization of a myoblast cell surface antigen
RT defined by 24.1B5 monoclonal antibody.";
RL Development 105:723-731(1993).
[3]
RN SEQUENCE OF 48-68 AND 213-246.
RX MEDLINE=93176127; PubMed=8439294;
RA Zhang W.-W., Zhang L.-X., Busch R.K., Farres J., Busch H.;
RT "Purification and characterization of a DNA-binding heterodimer of 52
RT and 100 kDa from HeLa cells.";
RL Biochem. J. 290:267-272(1993).

-- FUNCTION: ESSENTIAL PRE-MRNA SPLICING FACTOR REQUIRED EARLY IN
-- SPLICOSOME FORMATION. BINDS TO THE MAMMALIAN POLYPYRIMIDINE
-- TRACTS. FORMS A COMPLEX WITH THE POLYPYRIMIDINE TRACT-BINDING
-- PROTEIN (PTB). SEEMS TO ALSO BIND DNA.
-- SUBUNIT: HETEROTETRAMER OF TWO 52 KDA AND TWO 100 KDA SUBUNITS.
-- SUBCELLULAR LOCATION: Nuclear.
-- ALTERNATIVE PRODUCTS:
-- Event=Alternative splicing; Named isoforms=2;
-- Comment=Additional isoforms seem to exist;
-- Name=Long;
-- IsoIda=P23246-1; Sequence=Displayed;
-- Name=Short;
-- IsoIda=P23246-2; Sequence=VSP_005855;
-- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.
-- CAUTION: WAS ORIGINALLY (REF.2) THOUGHT TO BE MYOBLAST CELL
-- SURFACE ANTIGEN 24.1D5 AND A POSSIBLE MEMBRANE-BOUND PROTEIN
-- ECTONINASE.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)

EMBL; X70944; CAA50283.1; -
EMBL; X16850; CAA34747.1; -
PIR; A46302; A46302.
HSP; P11940; 1CVJ.
SWISS-2DPAGE; P23246; HUMAN.
Genew; HGNC:10774; SPFQ.
CK; P23246; -
MIM; 605139; -
GO; GO:0006248; F:pre-mRNA splicing factor activity; TAS.
GO; GO:0006371; P:mRNA splicing; TAS.
InterPro; IPR000504; RNA_rec_mot.
Pfam; PF00076; rrm; 2.
SMART; SM00360; RRM; 2.
PROSITE; PS01012; RRM; 2.
Nuclear protein; RNA-binding; DNA-binding; mRNA splicing; Repeat;
Alternative splicing.
FT DOMAIN 297 369 RNA-BINDING (RRM) 1.
FT DOMAIN 371 452 RNA-BINDING (RRM) 2.
FT DOMAIN 9 27 3 X 3 AA REPEATS OF R-G-G.
FT REPEAT 9 11 1.
FT REPEAT 19 21 2.
FT REPEAT 25 27 3.
FT DOMAIN 10 266 GLN/GLU/PRO-RICH.
FT DOMAIN 10 15 POLY-GLY.
FT DOMAIN 20 27 POLY-GLY.
FT DOMAIN 56 65 POLY-PRO.
FT DOMAIN 67 71 POLY-GLN.
FT DOMAIN 95 98 POLY-GLN.
FT DOMAIN 99 103 POLY-PRO.
FT DOMAIN 184 188 POLY-PRO.
FT DOMAIN 571 574 POLY-ARG.
FT DOMAIN 613 616 POLY-GLY.
FT DOMAIN 635 641 POLY-GLY.
FT VARSPIC 663 707
RTERFGCGAGPVGQGGPQGMGPGTPTAGTGRGREGVEGPNK
KPRF -> VMIDVG (in isoform Short).
/FTID=VSP_005855.
G -> R (IN REF. 3).

Query Match 16.3%; Score 167.5; DB 1; Length 707;
Best Local Similarity 40.0%; Pred. No. 0.031;
Matches 54; Conservative 5; Mismatches 47; Indels 29; Gaps 8;

QY 30 RNDPKLPQPPANQDQSSQNTRELQPTPAPAPK-----AAPPLDRESGPGV 83
Db 53 QSGKPK-FIPPPPHQCCQPPPPQPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP 111

```

QY 84 EN-KLIPSVGS--PASSTPLP-----PDGTGPNSTPN-NRAVTVSQGNSSSADP 130  
 Db 112 AGQGFAGVGAPPASSAPPATFTTSCAGPFGPGTTPPPAVT-----SAPP 162  
 QY 131 KAPPPPPVSSGEPPT 145  
 Db 163 GAPPPTPSSGVPTT 177

RESULT 3  
 MK07 HUMAN  
 ID MK07 HUMAN STANDARD; PRT; 815 AA.  
 AC Q13164; Q16634;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Mitogen-activated protein kinase 7 (EC 2.7.1.-) (Extracellular signal-  
 regulated kinase 5) (ERK-5) (ERK4) (BMK1 kinase).  
 GN MAPK7 OR PRKM7 OR ERK5 OR ERK4.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=fetal brain;  
 RX MEDLINE=95279403; PubMed=7759517;  
 RA Zhou G., Bao Z.Q., Dixon J.E.;  
 RT "Components of a new human protein kinase signal transduction  
 pathway.";  
 RL J. Biol. Chem. 270:12665-12669 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=95374539; PubMed=7646528;  
 RA Lee J.-D., Ulevitch R.J., Han J.;  
 RT "Primary structure of BMK1: a new mammalian map kinase.";  
 RL Biochem. Biophys. Res. Commun. 213:715-724 (1995).  
 CC -!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND  
 CC NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS.  
 CC -!- ENZYME REGULATION: Activated by tyrosine and threonine  
 CC phosphorylation (By similarity).  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY ADULT TISSUES. ABUNDANT IN  
 CC HEART, PLACENTA, LUNG, KIDNEY AND SKELETAL MUSCLE. NOT DETECTABLE  
 CC IN LIVER.  
 CC -!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN  
 CC TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.  
 CC -!- FTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN  
 CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY  
 CC ROLE, IS ABSENT.  
 CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
 CC MAP KINASE SUBFAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U25278; AAA81381.1;  
 DR EMBL; U29725; AAA82931.1;  
 DR EMBL; U29726; AAA82932.1;  
 DR EMBL; U29727; AAA82933.1;  
 DR EMBL; U29728; AAA82934.1;  
 DR EMBL; U29729; AAA82935.1;  
 DR EMBL; U29730; AAA82936.1;  
 DR EMBL; U29731; AAA82937.1;  
 DR EMBL; U29732; AAA82938.1;  
 DR EMBL; U29733; AAA82939.1;  
 DR EMBL; U29734; AAA82940.1;  
 DR EMBL; U29735; AAA82941.1;  
 DR EMBL; U29736; AAA82942.1;  
 DR EMBL; U29737; AAA82943.1;  
 DR EMBL; U29738; AAA82944.1;  
 DR EMBL; U29739; AAA82945.1;  
 DR EMBL; U29740; AAA82946.1;  
 DR EMBL; U29741; AAA82947.1;  
 DR EMBL; U29742; AAA82948.1;  
 DR EMBL; U29743; AAA82949.1;  
 DR EMBL; U29744; AAA82950.1;  
 DR EMBL; U29745; AAA82951.1;  
 DR EMBL; U29746; AAA82952.1;  
 DR EMBL; U29747; AAA82953.1;  
 DR EMBL; U29748; AAA82954.1;  
 DR EMBL; U29749; AAA82955.1;  
 DR EMBL; U29750; AAA82956.1;  
 DR EMBL; U29751; AAA82957.1;  
 DR EMBL; U29752; AAA82958.1;  
 DR EMBL; U29753; AAA82959.1;  
 DR EMBL; U29754; AAA82960.1;  
 DR EMBL; U29755; AAA82961.1;  
 DR EMBL; U29756; AAA82962.1;  
 DR EMBL; U29757; AAA82963.1;  
 DR EMBL; U29758; AAA82964.1;  
 DR EMBL; U29759; AAA82965.1;  
 DR EMBL; U29760; AAA82966.1;  
 DR EMBL; U29761; AAA82967.1;  
 DR EMBL; U29762; AAA82968.1;  
 DR EMBL; U29763; AAA82969.1;  
 DR EMBL; U29764; AAA82970.1;  
 DR EMBL; U29765; AAA82971.1;  
 DR EMBL; U29766; AAA82972.1;  
 DR EMBL; U29767; AAA82973.1;  
 DR EMBL; U29768; AAA82974.1;  
 DR EMBL; U29769; AAA82975.1;  
 DR EMBL; U29770; AAA82976.1;  
 DR EMBL; U29771; AAA82977.1;  
 DR EMBL; U29772; AAA82978.1;  
 DR EMBL; U29773; AAA82979.1;  
 DR EMBL; U29774; AAA82980.1;  
 DR EMBL; U29775; AAA82981.1;  
 DR EMBL; U29776; AAA82982.1;  
 DR EMBL; U29777; AAA82983.1;  
 DR EMBL; U29778; AAA82984.1;  
 DR EMBL; U29779; AAA82985.1;  
 DR EMBL; U29780; AAA82986.1;  
 DR EMBL; U29781; AAA82987.1;  
 DR EMBL; U29782; AAA82988.1;  
 DR EMBL; U29783; AAA82989.1;  
 DR EMBL; U29784; AAA82990.1;  
 DR EMBL; U29785; AAA82991.1;  
 DR EMBL; U29786; AAA82992.1;  
 DR EMBL; U29787; AAA82993.1;  
 DR EMBL; U29788; AAA82994.1;  
 DR EMBL; U29789; AAA82995.1;  
 DR EMBL; U29790; AAA82996.1;  
 DR EMBL; U29791; AAA82997.1;  
 DR EMBL; U29792; AAA82998.1;  
 DR EMBL; U29793; AAA82999.1;  
 DR EMBL; U29794; AAA83000.1;  
 DR EMBL; U29795; AAA83001.1;  
 DR EMBL; U29796; AAA83002.1;  
 DR EMBL; U29797; AAA83003.1;  
 DR EMBL; U29798; AAA83004.1;  
 DR EMBL; U29799; AAA83005.1;  
 DR EMBL; U29800; AAA83006.1;  
 DR EMBL; U29801; AAA83007.1;  
 DR EMBL; U29802; AAA83008.1;  
 DR EMBL; U29803; AAA83009.1;  
 DR EMBL; U29804; AAA83010.1;  
 DR EMBL; U29805; AAA83011.1;  
 DR EMBL; U29806; AAA83012.1;  
 DR EMBL; U29807; AAA83013.1;  
 DR EMBL; U29808; AAA83014.1;  
 DR EMBL; U29809; AAA83015.1;  
 DR EMBL; U29810; AAA83016.1;  
 DR EMBL; U29811; AAA83017.1;  
 DR EMBL; U29812; AAA83018.1;  
 DR EMBL; U29813; AAA83019.1;  
 DR EMBL; U29814; AAA83020.1;  
 DR EMBL; U29815; AAA83021.1;  
 DR EMBL; U29816; AAA83022.1;  
 DR EMBL; U29817; AAA83023.1;  
 DR EMBL; U29818; AAA83024.1;  
 DR EMBL; U29819; AAA83025.1;  
 DR EMBL; U29820; AAA83026.1;  
 DR EMBL; U29821; AAA83027.1;  
 DR EMBL; U29822; AAA83028.1;  
 DR EMBL; U29823; AAA83029.1;  
 DR EMBL; U29824; AAA83030.1;  
 DR EMBL; U29825; AAA83031.1;  
 DR EMBL; U29826; AAA83032.1;  
 DR EMBL; U29827; AAA83033.1;  
 DR EMBL; U29828; AAA83034.1;  
 DR EMBL; U29829; AAA83035.1;  
 DR EMBL; U29830; AAA83036.1;  
 DR EMBL; U29831; AAA83037.1;  
 DR EMBL; U29832; AAA83038.1;  
 DR EMBL; U29833; AAA83039.1;  
 DR EMBL; U29834; AAA83040.1;  
 DR EMBL; U29835; AAA83041.1;  
 DR EMBL; U29836; AAA83042.1;  
 DR EMBL; U29837; AAA83043.1;  
 DR EMBL; U29838; AAA83044.1;  
 DR EMBL; U29839; AAA83045.1;  
 DR EMBL; U29840; AAA83046.1;  
 DR EMBL; U29841; AAA83047.1;  
 DR EMBL; U29842; AAA83048.1;  
 DR EMBL; U29843; AAA83049.1;  
 DR EMBL; U29844; AAA83050.1;  
 DR EMBL; U29845; AAA83051.1;  
 DR EMBL; U29846; AAA83052.1;  
 DR EMBL; U29847; AAA83053.1;  
 DR EMBL; U29848; AAA83054.1;  
 DR EMBL; U29849; AAA83055.1;  
 DR EMBL; U29850; AAA83056.1;  
 DR EMBL; U29851; AAA83057.1;  
 DR EMBL; U29852; AAA83058.1;  
 DR EMBL; U29853; AAA83059.1;  
 DR EMBL; U29854; AAA83060.1;  
 DR EMBL; U29855; AAA83061.1;  
 DR EMBL; U29856; AAA83062.1;  
 DR EMBL; U29857; AAA83063.1;  
 DR EMBL; U29858; AAA83064.1;  
 DR EMBL; U29859; AAA83065.1;  
 DR EMBL; U29860; AAA83066.1;  
 DR EMBL; U29861; AAA83067.1;  
 DR EMBL; U29862; AAA83068.1;  
 DR EMBL; U29863; AAA83069.1;  
 DR EMBL; U29864; AAA83070.1;  
 DR EMBL; U29865; AAA83071.1;  
 DR EMBL; U29866; AAA83072.1;  
 DR EMBL; U29867; AAA83073.1;  
 DR EMBL; U29868; AAA83074.1;  
 DR EMBL; U29869; AAA83075.1;  
 DR EMBL; U29870; AAA83076.1;  
 DR EMBL; U29871; AAA83077.1;  
 DR EMBL; U29872; AAA83078.1;  
 DR EMBL; U29873; AAA83079.1;  
 DR EMBL; U29874; AAA83080.1;  
 DR EMBL; U29875; AAA83081.1;  
 DR EMBL; U29876; AAA83082.1;  
 DR EMBL; U29877; AAA83083.1;  
 DR EMBL; U29878; AAA83084.1;  
 DR EMBL; U29879; AAA83085.1;  
 DR EMBL; U29880; AAA83086.1;  
 DR EMBL; U29881; AAA83087.1;  
 DR EMBL; U29882; AAA83088.1;  
 DR EMBL; U29883; AAA83089.1;  
 DR EMBL; U29884; AAA83090.1;  
 DR EMBL; U29885; AAA83091.1;  
 DR EMBL; U29886; AAA83092.1;  
 DR EMBL; U29887; AAA83093.1;  
 DR EMBL; U29888; AAA83094.1;  
 DR EMBL; U29889; AAA83095.1;  
 DR EMBL; U29890; AAA83096.1;  
 DR EMBL; U29891; AAA83097.1;  
 DR EMBL; U29892; AAA83098.1;  
 DR EMBL; U29893; AAA83099.1;  
 DR EMBL; U29894; AAA83100.1;  
 DR EMBL; U29895; AAA83101.1;  
 DR EMBL; U29896; AAA83102.1;  
 DR EMBL; U29897; AAA83103.1;  
 DR EMBL; U29898; AAA83104.1;  
 DR EMBL; U29899; AAA83105.1;  
 DR EMBL; U29900; AAA83106.1;  
 DR EMBL; U29901; AAA83107.1;  
 DR EMBL; U29902; AAA83108.1;  
 DR EMBL; U29903; AAA83109.1;  
 DR EMBL; U29904; AAA83110.1;  
 DR EMBL; U29905; AAA83111.1;  
 DR EMBL; U29906; AAA83112.1;  
 DR EMBL; U29907; AAA83113.1;  
 DR EMBL; U29908; AAA83114.1;  
 DR EMBL; U29909; AAA83115.1;  
 DR EMBL; U29910; AAA83116.1;  
 DR EMBL; U29911; AAA83117.1;  
 DR EMBL; U29912; AAA83118.1;  
 DR EMBL; U29913; AAA83119.1;  
 DR EMBL; U29914; AAA83120.1;  
 DR EMBL; U29915; AAA83121.1;  
 DR EMBL; U29916; AAA83122.1;  
 DR EMBL; U29917; AAA83123.1;  
 DR EMBL; U29918; AAA83124.1;  
 DR EMBL; U29919; AAA83125.1;  
 DR EMBL; U29920; AAA83126.1;  
 DR EMBL; U29921; AAA83127.1;  
 DR EMBL; U29922; AAA83128.1;  
 DR EMBL; U29923; AAA83129.1;  
 DR EMBL; U29924; AAA83130.1;  
 DR EMBL; U29925; AAA83131.1;  
 DR EMBL; U29926; AAA83132.1;  
 DR EMBL; U29927; AAA83133.1;  
 DR EMBL; U29928; AAA83134.1;  
 DR EMBL; U29929; AAA83135.1;  
 DR EMBL; U29930; AAA83136.1;  
 DR EMBL; U29931; AAA83137.1;  
 DR EMBL; U29932; AAA83138.1;  
 DR EMBL; U29933; AAA83139.1;  
 DR EMBL; U29934; AAA83140.1;  
 DR EMBL; U29935; AAA83141.1;  
 DR EMBL; U29936; AAA83142.1;  
 DR EMBL; U29937; AAA83143.1;  
 DR EMBL; U29938; AAA83144.1;  
 DR EMBL; U29939; AAA83145.1;  
 DR EMBL; U29940; AAA83146.1;  
 DR EMBL; U29941; AAA83147.1;  
 DR EMBL; U29942; AAA83148.1;  
 DR EMBL; U29943; AAA83149.1;  
 DR EMBL; U29944; AAA83150.1;  
 DR EMBL; U29945; AAA83151.1;  
 DR EMBL; U29946; AAA83152.1;  
 DR EMBL; U29947; AAA83153.1;  
 DR EMBL; U29948; AAA83154.1;  
 DR EMBL; U29949; AAA83155.1;  
 DR EMBL; U29950; AAA83156.1;  
 DR EMBL; U29951; AAA83157.1;  
 DR EMBL; U29952; AAA83158.1;  
 DR EMBL; U29953; AAA83159.1;  
 DR EMBL; U29954; AAA83160.1;  
 DR EMBL; U29955; AAA83161.1;  
 DR EMBL; U29956; AAA83162.1;  
 DR EMBL; U29957; AAA83163.1;  
 DR EMBL; U29958; AAA83164.1;  
 DR EMBL; U29959; AAA83165.1;  
 DR EMBL; U29960; AAA83166.1;  
 DR EMBL; U29961; AAA83167.1;  
 DR EMBL; U29962; AAA83168.1;  
 DR EMBL; U29963; AAA83169.1;  
 DR EMBL; U29964; AAA83170.1;  
 DR EMBL; U29965; AAA83171.1;  
 DR EMBL; U29966; AAA83172.1;  
 DR EMBL; U29967; AAA83173.1;  
 DR EMBL; U29968; AAA83174.1;  
 DR EMBL; U29969; AAA83175.1;  
 DR EMBL; U29970; AAA83176.1;  
 DR EMBL; U29971; AAA83177.1;  
 DR EMBL; U29972; AAA83178.1;  
 DR EMBL; U29973; AAA83179.1;  
 DR EMBL; U29974; AAA83180.1;  
 DR EMBL; U29975; AAA83181.1;  
 DR EMBL; U29976; AAA83182.1;  
 DR EMBL; U29977; AAA83183.1;  
 DR EMBL; U29978; AAA83184.1;  
 DR EMBL; U29979; AAA83185.1;  
 DR EMBL; U29980; AAA83186.1;  
 DR EMBL; U29981; AAA83187.1;  
 DR EMBL; U29982; AAA83188.1;  
 DR EMBL; U29983; AAA83189.1;  
 DR EMBL; U29984; AAA83190.1;  
 DR EMBL; U29985; AAA83191.1;  
 DR EMBL; U29986; AAA83192.1;  
 DR EMBL; U29987; AAA83193.1;  
 DR EMBL; U29988; AAA83194.1;  
 DR EMBL; U29989; AAA83195.1;  
 DR EMBL; U29990; AAA83196.1;  
 DR EMBL; U29991; AAA83197.1;  
 DR EMBL; U29992; AAA83198.1;  
 DR EMBL; U29993; AAA83199.1;  
 DR EMBL; U29994; AAA83200.1;  
 DR EMBL; U29995; AAA83201.1;  
 DR EMBL; U29996; AAA83202.1;  
 DR EMBL; U29997; AAA83203.1;  
 DR EMBL; U29998; AAA83204.1;  
 DR EMBL; U29999; AAA83205.1;  
 DR EMBL; U30000; AAA83206.1;  
 DR EMBL; U30001; AAA83207.1;  
 DR EMBL; U30002; AAA83208.1;  
 DR EMBL; U30003; AAA83209.1;  
 DR EMBL; U30004; AAA83210.1;  
 DR EMBL; U30005; AAA83211.1;  
 DR EMBL; U30006; AAA83212.1;  
 DR EMBL; U30007; AAA83213.1;  
 DR EMBL; U30008; AAA83214.1;  
 DR EMBL; U30009; AAA83215.1;  
 DR EMBL; U30010; AAA83216.1;  
 DR EMBL; U30011; AAA83217.1;  
 DR EMBL; U30012; AAA83218.1;  
 DR EMBL; U30013; AAA83219.1;  
 DR EMBL; U30014; AAA83220.1;  
 DR EMBL; U30015; AAA83221.1;  
 DR EMBL; U30016; AAA83222.1;  
 DR EMBL; U30017; AAA83223.1;  
 DR EMBL; U30018; AAA83224.1;  
 DR EMBL; U30019; AAA83225.1;  
 DR EMBL; U30020; AAA83226.1;  
 DR EMBL; U30021; AAA83227.1;  
 DR EMBL; U30022; AAA83228.1;  
 DR EMBL; U30023; AAA83229.1;  
 DR EMBL; U30024; AAA83230.1;  
 DR EMBL; U30025; AAA83231.1;  
 DR EMBL; U30026; AAA83232.1;  
 DR EMBL; U30027; AAA83233.1;  
 DR EMBL; U30028; AAA83234.1;  
 DR EMBL; U30029; AAA83235.1;  
 DR EMBL; U30030; AAA83236.1;  
 DR EMBL; U30031; AAA83237.1;  
 DR EMBL; U30032; AAA83238.1;  
 DR EMBL; U30033; AAA83239.1;  
 DR EMBL; U30034; AAA83240.1;  
 DR EMBL; U30035; AAA83241.1;  
 DR EMBL; U30036; AAA83242.1;  
 DR EMBL; U30037; AAA83243.1;  
 DR EMBL; U30038; AAA83244.1;  
 DR EMBL; U30039; AAA83245.1;  
 DR EMBL; U30040; AAA83246.1;  
 DR EMBL; U30041; AAA83247.1;  
 DR EMBL; U30042; AAA83248.1;  
 DR EMBL; U30043; AAA83249.1;  
 DR EMBL; U30044; AAA83250.1;  
 DR EMBL; U30045; AAA83251.1;  
 DR EMBL; U30046; AAA83252.1;  
 DR EMBL; U30047; AAA83253.1;  
 DR EMBL; U30048; AAA83254.1;  
 DR EMBL; U30049; AAA83255.1;  
 DR EMBL; U30050; AAA83256.1;  
 DR EMBL; U30051; AAA83257.1;  
 DR EMBL; U30052; AAA83258.1;  
 DR EMBL; U30053; AAA83259.1;  
 DR EMBL; U30054; AAA83260.1;  
 DR EMBL; U30055; AAA83261.1;  
 DR EMBL; U30056; AAA83262.1;  
 DR EMBL; U30057; AAA83263.1;  
 DR EMBL; U30058; AAA83264.1;  
 DR EMBL; U30059; AAA83265.1;  
 DR EMBL; U30060; AAA83266.1;  
 DR EMBL; U30061; AAA83267.1;  
 DR EMBL; U30062; AAA83268.1;  
 DR EMBL; U30063; AAA83269.1;  
 DR EMBL; U30064; AAA83270.1;  
 DR EMBL; U30065; AAA83271.1;  
 DR EMBL; U30066; AAA83272.1;  
 DR EMBL; U30067; AAA83273.1;  
 DR EMBL; U30068; AAA83274.1;  
 DR EMBL; U30069; AAA83275.1;  
 DR EMBL; U30070; AAA83276.1;  
 DR EMBL; U30071; AAA83277.1;  
 DR EMBL; U30072; AAA83278.1;  
 DR EMBL; U30073; AAA83279.1;  
 DR EMBL; U30074; AAA83280.1;  
 DR EMBL; U30075; AAA83281.1;  
 DR EMBL; U30076; AAA83282.1;  
 DR EMBL; U30077; AAA83283.1;  
 DR EMBL; U30078; AAA83284.1;  
 DR EMBL; U30079; AAA83285.1;  
 DR EMBL; U30080; AAA83286.1;  
 DR EMBL; U30081; AAA83287.1;  
 DR EMBL; U30082; AAA83288.1;  
 DR EMBL; U30083; AAA83289.1;  
 DR EMBL; U30084; AAA83290.1;  
 DR EMBL; U30085; AAA83291.1;  
 DR EMBL; U30086; AAA83292.1;  
 DR EMBL; U30087; AAA83293.1;  
 DR EMBL; U30088; AAA83294.1;  
 DR EMBL; U30089; AAA83295.1;  
 DR EMBL; U30090; AAA83296.1;  
 DR EMBL; U30091; AAA83297.1;  
 DR EMBL; U30092; AAA83298.1;  
 DR EMBL; U30093; AAA83299.1;  
 DR EMBL; U30094; AAA83300.1;  
 DR EMBL; U30095; AAA83301.1;  
 DR EMBL; U30096; AAA83302.1;  
 DR EMBL; U30097; AAA83303.1;  
 DR EMBL; U30098; AAA83304.1;  
 DR EMBL; U30099; AAA83305.1;  
 DR EMBL; U30100; AAA83306.1;  
 DR EMBL; U30101; AAA83307.1;  
 DR EMBL; U30102; AAA83308.1;  
 DR EMBL; U30103; AAA83309.1;  
 DR EMBL; U30104; AAA83310.1;  
 DR EMBL; U30105; AAA83311.1;  
 DR EMBL; U30106; AAA83312.1;  
 DR EMBL; U30107; AAA83313.1;  
 DR EMBL; U30108; AAA83314.1;  
 DR EMBL; U30109; AAA83315.1;  
 DR EMBL; U30110; AAA83316.1;  
 DR EMBL; U30111; AAA83317.1;  
 DR EMBL; U30112; AAA83318.1;  
 DR EMBL; U30113; AAA83319.1;  
 DR EMBL; U30114; AAA83320.1;  
 DR EMBL; U30115; AAA83321.1;  
 DR EMBL; U30116; AAA83322.1;  
 DR EMBL; U30117; AAA83323.1;  
 DR EMBL; U30118; AAA83324.1;  
 DR EMBL; U30119; AAA83325.1;  
 DR EMBL; U30120; AAA83326.1;  
 DR EMBL; U30121; AAA83327.1;  
 DR EMBL; U30122; AAA83328.1;  
 DR EMBL; U30123; AAA83329.1;  
 DR EMBL; U30124; AAA83330.1;  
 DR EMBL; U30125; AAA83331.1;  
 DR EMBL; U30126; AAA83332.1;  
 DR EMBL; U30127; AAA83333.1;  
 DR EMBL; U30128; AAA83334.1;  
 DR EMBL; U30129; AAA83335.1;  
 DR EMBL; U30130; AAA83336.1;  
 DR EMBL; U30131; AAA83337.1;  
 DR EMBL; U30132; AAA83338.1;  
 DR EMBL; U30133; AAA83339.1;  
 DR EMBL; U30134; AAA83340.1;  
 DR EMBL; U30135; AAA83341.1;  
 DR EMBL; U30136; AAA83342.1;  
 DR EMBL; U30137; AAA83343.1;  
 DR EMBL; U30138; AAA83344.1;  
 DR EMBL; U30139; AAA83345.1;  
 DR EMBL; U30140; AAA83346.1;  
 DR EMBL; U30141; AAA83347.1;  
 DR EMBL; U30142; AAA83348.1;  
 DR EMBL; U30143; AAA83349.1;  
 DR EMBL; U30144; AAA83350.1;  
 DR EMBL; U30145; AAA83351.1;  
 DR EMBL; U30146; AAA83352.1;  
 DR EMBL; U30147; AAA83353.1;  
 DR EMBL; U30148; AAA83354.1;  
 DR EMBL; U30149; AAA83355.1;  
 DR EMBL; U30150; AAA83356.1;  
 DR EMBL; U30151; AAA83357.1;  
 DR EMBL; U30152; AAA83358.1;  
 DR EMBL; U30153; AAA83359.1;  
 DR EMBL; U30154; AAA83360.1;  
 DR EMBL; U30155; AAA83361.1;  
 DR EMBL; U30156; AAA83362.1;  
 DR EMBL; U30157; AAA83363.1;  
 DR EMBL; U30158; AAA83364.1;  
 DR EMBL; U30159; AAA83365.1;  
 DR EMBL; U30160; AAA83366.1;  
 DR EMBL; U30161; AAA83367.1;  
 DR EMBL; U30162; AAA83368.1;  
 DR EMBL; U30163; AAA83369.1;  
 DR EMBL; U30164; AAA83370.1;  
 DR EMBL; U30165; AAA

```

Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
-!- FUNCTION: Major component of the outer cell wall w6 (crystalline)
layer.
-!- SUBUNIT: Associates with GP2 and GP3.
-!- PM: N-glycosylated and O-glycosylated.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF108494; M645420.1; ALT_SEQ.
EMBL; M58496; AAG59706.1; ALT_SEQ.
Glycosylated; O-glycosylated.
InterPro: IPR002365; P-rich extensin.
InterPro: IPR003882; P-rich extensin.
PRINTS; PRO1217; PRICHEXTENSIN.
DR; PRO1218; PSTLEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GPI.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSPX REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 399 399 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;
-----
Query Match 15.5%; Score 159.5; DB 1; Length 555;
Best Local Similarity 32.8%; Pred. No. 0.063;
Matches 38; Conservative 6; Mismatches 49; Indels 23; Gaps 3;
-----
QY 33 PKPLQCPQAPANQDQNSQNTRLQPTLP-----APAPPAAPRLDLSRGVKNLI 88
DB 179 PPSPPVPSPAPP-----SFAPVPPSPAPPSPAPPVPPSPAPPSPAPPSP 228
QY 89 PSVGSPPASLTPLDCTGNTGNVAVTFVSCQNSGSSADPKAPPPPPVSSGEPP 144
DB 229 PGPAPPSPSPAPPSPVPVPPSPAPPSPA-----PPSPKPPAPPSPPPPPPP 275
-----
RESULT 5
ENAH_MOUSE
ID ENAH_MOUSE STANDARD; PRT; 802 AA.
AC Q03173; P70430; P70431; P70432; P70433;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Enabled protein homolog (NFC derived proline-rich protein 1) (NDPP-1).
GN ENAH OR MENA OR NDPP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Sakuma T., Tomooka Y., Katjhu S., Ikawa Y., Noda M., Kumar S.;
RT "Identification of a developmentally regulated gene in the mouse
central nervous system which encodes a novel proline rich protein.";
RL Biochim. Biophys. Acta 113:240-248(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4 AND 5), FUNCTION, AND SUBCELLULAR
LOCATION.
RC TISSUE=Brain;
RA Gertler F.B., Nebuhr K., Reinhard M., Wehland J., Soriano P.;
RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
control of microfilament dynamics.";
-----
Cell 87:227-239 (1996).
[3]
RN FUNCTION, AND SUBUNIT.
RP MEDLINE-9916867; PubMed=10069337; Menzies A.S., Wehman A.M.,
RA Lanier L.M., Gates M.A., Mitke W., Koziakowski D., Soriano P., Gertler F.B.;
RT "Mena is required for neurulation and commissure formation.";
RL Neuron 22:313-325 (1999).
CC -!- FUNCTION: May be involved in microfilament assembly and cell
motility. Induces the formation of F-actin rich outgrowths in
microtubules. Required for neurulation and commissure formation.
CC -!- SUBUNIT: Binds profilin.
CC -!- SUBCELLULAR LOCATION: Localized to focal adhesions and, to a
lesser extent, leading edges and stress fibers.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=5; Synonyms=Mena++;
CC ISOID=Q03173-1; Sequence=Displayed;
CC Name=1;
CC ISOID=Q03173-2; Sequence=VSP_007255; VSP_007256;
CC Name=2; Synonyms=Mena;
CC ISOID=Q03173-3; Sequence=VSP_007259; VSP_007260;
CC Name=3; Synonyms=Mena+;
CC ISOID=Q03173-4; Sequence=VSP_007259;
CC Name=4; Synonyms=Mena++;
CC ISOID=Q03173-5; Sequence=VSP_007257, VSP_007258;
CC -!- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
CC -!- SIMILARITY: Contains 1 WH1 domain.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; D10727; BAA01570.1; -
EMBL; U72520; AAC52863.1; -
EMBL; U72521; AAC52864.1; -
EMBL; U72522; AAC52865.1; -
EMBL; U72523; AAC52866.1; -
DR PIR; S27200; S27200.
DR MGI; MGI:108360; Enah.
DR InterPro; IPR001960; WH1.
DR Pfam; PF00568; WH1; 1.
DR SMART; SM00461; WH1; 1.
KW Developmental protein; Neurogenesis; Alternative splicing.
FT DOMAIN 1 108 WH1
FT DOMAIN 442 464 POLY-PRO.
FT DOMAIN 542 552 POLY-PRO.
FT DOMAIN 562 574 POLY-PRO.
FT DOMAIN 578 589 POLY-PRO.
FT DOMAIN 593 605 POLY-PRO.
FT VARSPPLIC 1 412 Missing (in isoform 1).
FT VARSPPLIC 535 631 /FTID=VSP_007255.
FT TQGLVLGPPAPPPPPPLPSGPAYASALPPPPPPPPPLPS
FT TGPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
FT DNRPLTGLAAAIAAG -> HRAASDHLHRLHRLHSHQALP
FT TPQHFLLQDPLHLHLHCHPLVLLHPLHPLHPLKLPPLP
FT HLLPLSPHLEFSLDRQKTIANFTWCNCGS (in
FT isoform 1).
FT /FTID=VSP_007256.
FT Missing (in isoform 4).
FT /FTID=VSP_007257.
FT CIPC -> VFYL (in isoform 4).
FT /FTID=VSP_007258.
FT Missing (in isoform 2 and isoform 3).
FT /FTID=VSP_007259.
FT Missing (in isoform 2).
FT /FTID=VSP_007260.
FT /FTID=VSP_007260.
SQ SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;
-----

```

Matches	39;	Conservative	9;	Mismatches	57;	Indels	12;	Gaps	3;
---------	-----	--------------	----	------------	-----	--------	-----	------	----

  

QY	33	KPELPQPPAPAWQDQNSQNTLRQLTPTIPAPAPKAPAPRPDLDRSPGVENKLI	PSVVG	92
DB	2	KPELPQPPAPAWQDQNSQNTLRQLTPTIPAPAPKAPAPRPDLDRSPGVENKLI	PSVVG	92
QY	93	-----SPASSPTPLPPDGTGTPNSFNNAVTVSCGNSSSADPKAPPPVPSSGEPP	144	
DB	62	PSVKGFGSGFPPSPAPKVPD-----PSSSPKPSPPKPPSPSPR-SPKPP	111	

  

RESULT 7				
WAIP	HUMAN	STANDARD;	PRT;	503 AA.
AC	043516;	Q15220;	OSUNP1;	
DT	16-OCT-2001	(Rel. 40, Created)		
DT	16-OCT-2001	(Rel. 40, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Wiskott-Aldrich syndrome protein interacting protein (WASP interacting protein) (PRLP-2 protein).			
DE	WASP/PI OR WIP.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Eutelestomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxId:9606;			
RP	[1]	SEQUENCE FROM N.A.		
RP	MDLINE=98070810; PubMed=9405671;			
RA	Ramesh N, Anton I M, Hartwig J H, Geha R S ;			
RA	"Wip, a protein associated with Wiskott-Aldrich syndrome protein,			
RT	induces actin polymerization and redistribution in lymphoid cells,"			
EL	Proc. Natl. Acad. Sci. U.S.A. 94:14671-14676(1997).			
RP	[2]	SEQUENCE FROM N.A.		
RP	TISSUE=tonsil;			
NC	Kreidewitz S., Delany-Heiken P., Nordheim A., Ruhlmann A.;			
NC	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.			
RP	[3]	SEQUENCE OF 192-503 FROM N.A., AND VARIANT ALA-495.		
MD	MDLINE=99218549; PubMed=10202051;			
RA	Stewart D.M., Tian L., Nelson D.L.;			
RT	"Mutations that cause the Wiskott-Aldrich syndrome impair the			
RT	interaction of Wiskott-Aldrich syndrome protein (WASP) with WASP			
RT	interacting protein."			
RP	J. Immunol. 162:5019-5024(1999).			
CC	-!- FUNCTION: MAY HAVE DIRECT ACTIVITY ON THE ACTIN CYTOSKELETON.			
CC	-!- INDUCES ACTIN POLYMERIZATION AND REDISTRIBUTION.			
CC	-!- SUBUNIT: BINDS TO WASP, PROFILIN AND ACTIN.			
CC	-!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PERIPHERAL BLOOD			
CC	MONONUCLEAR CELLS, SPLEEN, PLACENTA, SMALL INTESTIN, COLON,			
CC	THYMUS. LOWER EXPRESSION IN OVARY, HEART, BRAIN, LUNG, LIVER,			
CC	SKELETAL MUSCLE, KIDNEY, PANCREAS, PROSTATE, TESTIS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
EMBL	AF031586; AAC03767.1; -			
DR	EMBL; AF031586; AAC03767.1; -			
DR	EMBL; X86019; CNA60014.1; -			
DR	Genew; HGNC:12736; WASP1P.			
DR	MM; 602357; -			
DR	GO; GO:0015629; Cactin cytoskeleton; TAS.			
DR	GO; GO:0003779; F-actin binding activity; TAS.			
DR	GO; GO:0005552; Profilin binding activity; TAS.			
DR	GO; GO:0008154; Protein polymerization and/or depolymerization; TAS.			
DR	GO; GO:0006461; P:protein complex assembly; TAS.			
DR	InterPro; IPR003124; WH2.			
DR	Pfam; PF02205; WH2.1.			

DR SMART; SM00246; WH2; 1.  
 KW Actin-binding; Repeat; Polymorphism.  
 FT DOMAIN 2 13 POLY-PRO.  
 FT DOMAIN 64 96 GLY-RICH.  
 FT DOMAIN 241 244 POLY-RICH.  
 FT DOMAIN 264 433 PRO-RICH.  
 FT REPEAT 352 361 XRSQPPXP MOTIF 1.  
 FT REPEAT 374 383 XRSQPPXP MOTIF 2.  
 FT REPEAT 410 419 XRSQPPXP MOTIF 3.  
 FT SITE 45 48 ACTIN BINDING.  
 FT VARIANT 495 495 G -> A.  
 FT /FTID=VAR\_010295.  
 FT CONFLICT 303 309 PHRPHLR -> SSOAPP (IN REF. 3).  
 FT CONFLICT 360 360 P -> PV (IN REF. 2).  
 FT CONFLICT 487 503 GSNRRERGGPLPPIR -> EYEQGF (IN REF. 2).  
 SQ SEQUENCE 503 AA; 51489 MW; 43E88674DD3BF1A CRC64;

Query Match 14.7%; Score 151.5; DB 1; Length 503;  
 Best Local Similarity 26.5%; Pred. NO. 0.15; Indels 57; Gaps 11;  
 Matches 60; Conservative 25; Mismatches 84; Indels 57; Gaps 11;

QY 12 NKK-----TERSTAPLNTQISALRNDKPLFCQPPAPADQNS-----SQNTL-OPT 59  
 DB 289 NKKPVVSTPRSAHRRH---LRPPPSRPGPLPPLPSSGNDLPLRQNLISLSST 345  
 QY 60 PPIPAKPAAPRPLDRSPGVENKLPISVSPASSTPLPDGTGPNSTNNRAV---116  
 DB 346 PLLSPGRSGPLPFPSPRPVPRD-----PGRSGPLPPLPVPVSRNGSTRALPAT 398  
 QY 117 --TPVSGSNSSADPK-----APPPPVSGSPPTLGENPDGLSQEQLHR--161  
 DB 399 POLPSRSGVDSRSGRPLPDRPDRSAGAPPPLPSTSRNGFQDSP---CEDENESRY 455  
 QY 162 -----ERSLQTLRIQMLFPDEKFTGA---QSGGPOQNP 194  
 DB 456 FHPTSLDLPPEPVVQTTSYKSLANNSR-SGSNRRERGGLPLP 500

RESULT 8  
 ID\_PP2\_HUMAN STANDARD; PRT; 251 AA.  
 AC P02812;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Salivary proline-rich protein precursor (Clone C97) [Contains: Basic Peptide P-F] (Fragment).  
 GN PRB2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
 RX MEDLINE=85289325; PubMed=2993301;  
 RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
 RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";  
 RL J. Biol. Chem. 260:11123-11130(1985).  
 RN [2]  
 RP SEQUENCE OF 134-194.  
 RX MEDLINE=83265674; PubMed=6874669;  
 RA Saitoh E., Iemura S., Sanada K.;  
 RT "Complete amino acid sequence of a basic proline-rich peptide, P-F, from human parotid saliva";  
 RL J. Biochem. 93:883-886(1983).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; K03208; AAA60189.1; -  
 DR PIR; E25372; PIHUPF.  
 DR Genew; HGNC:9338; PRB2.  
 DR MIM; 168810; -  
 KW Repeat; Parotid gland; Multigene family.  
 FT NON\_TER 1 1  
 FT CHAIN 134 194 BASIC PEPTIDE P-F.  
 FT REPEAT <1 9  
 FT REPEAT 10 71  
 FT REPEAT 72 133  
 FT REPEAT 134 195  
 SQ SEQUENCE 251 AA; 24641 MW; D779F590C0EBF30B CRC64;

Query Match 14.6%; Score 150.5; DB 1; Length 251;  
 Best Local Similarity 31.2%; Pred. NO. 0.091;  
 Matches 43; Conservative 9; Mismatches 61; Indels 25; Gaps 5;

QY 31 NDPK--PLPQPPAPANQDQNSQNTLRQTTP-----IPAPAKPAAPRP 75  
 DB 45 NKPGPPPKPKPGPPPGQNGKSGARSPPKPGQPGQNGKPGPPPKPGQPGPPQ 104  
 QY 76 LDRESGVENKLPISVSPASSTPLPDGTGPNSTNNRAVTPVSGSNSSADPKAPP 135  
 DB 105 GDNKSQSP-----PPCKPGQPPPGQSGKSRSSRSPKPGQPGPPQGN----QPGQPPP 155  
 QY 136 PP-VSSGEPPTLGENPDG 152  
 DB 156 PPGKPGQPPPGKNGKPG 173

RESULT 9  
 ID\_PP2\_HUMAN STANDARD; PRT; 533 AA.  
 AC P28702; P28703;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Retinoic acid receptor RXR-beta.  
 GN RXR OR NR2B2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
 RX MEDLINE=92253386; PubMed=1315958;  
 RA Fleischhauer K., Park J.H., Disanto J.P., Marks M.S., Ozato K., Yang S.Y.;  
 RT "Isolation of a full-length cDNA clone encoding a N-terminally variant form of the human retinoid X receptor beta.";  
 RL Nucleic Acids Res. 20:1801-1801(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Breast Carcinoma;  
 RX MEDLINE=92127595; PubMed=1310259;  
 RA Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M., Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S., Chambon P.;  
 RT "Purification, cloning, and RXR identity of the HeLa cell factor with which RXR or TR heterodimerizes to bind target sequences efficiently.";  
 RL Cell 68:377-395(1992).  
 RN [3]  
 RP REVISIONS.  
 RX MEDLINE=93046692; PubMed=1330328;  
 RA Leid M., Kastner P., Lyons R., Nakshatri H., Saunders M., Zacharewski T., Chen J.Y., Staub A., Garnier J.M., Mader S., Chambon P.;  
 RT Cell 71:887-887(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.

RA Numasawa T., Koga H., Ueyama K., Maeda S., Sakou T., Harata S.,  
 RA Leppert M., Inoue T.,  
 RA Submitted (May-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RN SEQUENCE FROM N.A.  
 RA Corella A., Vergara A., Paaz G., de Miguel C., Encio I.:  
 RA "Molecular cloning and characterization of the human HXR9B gene and 5'  
 RT flanking region";  
 RL Submitted (Jan-1999) to the EMBL/GenBank/DBJ databases.  
 RL [6]  
 RN SEQUENCE FROM N.A.  
 RA Tubby B.:  
 RA Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 RL [7]  
 RN SEQUENCE FROM N.A.  
 RA TISSUE=Brain;  
 RC MEDLINE=2238257; PubMed=12477937;  
 RA Strausberg R.D., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Soak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Gorman K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield A.S., N. Krzywicki M.I., Skalska U., Smallos D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.:  
 RA "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences";  
 RL Proc. Natl Acad Sci U.S.A. 99:16899-16903 (2002).  
 CC -|- FUNCTION: INVOLVED IN RETINOIC ACID RESPONSE PATHWAY. BINDS 9-CIS  
 CC RETINOIC ACID (9C-RAR).  
 CC -|- SUBCELLULAR LOCATION: Nuclear.  
 CC -|- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Comment-Additional isoforms seem to exist;  
 CC Names=Long;  
 CC IsoId=P28702-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=P28702-2; Sequence=Not described;  
 CC -|- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -|- SIMILARITY: Belongs to the nuclear hormone receptor family. NR2  
 CC subfamily.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X63522; CAA45087.1; -;  
 CC EMBL; M84820; AAA60293.1; -;  
 CC EMBL; AF065596; AAC18599.1; -;  
 CC EMBL; AF120161; AAD13794.1; -;  
 CC EMBL; AL031228; CAA20239.1; -;  
 CC EMBL; BC001167; AAH01167.1; -;  
 CC FIB; S17781; S17781;  
 CC PDB; 1H9U; 22-MAY-02.  
 CC Genew; HGNC:10478; RXRB.  
 CC TRANSFAC; T01334; -;  
 CC MIM; 160246; -;  
 CC GO; GO:0004886; Fretinoid-X receptor activity; TAS.  
 CC GO; GO:0003713; Fretinoid-X receptor co-activator activity; TAS.

DE InterPro; IPR000536; Hormone\_rec\_1lg.  
 DE InterPro; IPR001723; Steroid\_receptor.  
 DE InterPro; IPR001628; Znf\_C4steroid.  
 DE Pfam; PF00104; hormone\_rec; 1.  
 DE Pfam; PF00105; zfc4; 1.  
 DE PRINTS; PR00398; STERHORMONER.  
 DE PRINTS; PR00047; STERDIFINGER.  
 DE Prodom; PD000035; Znf\_C4steroid; 1.  
 DE SMART; SM00430; ROLI; 1.  
 DE SMART; SM00399; Znf\_C4; 1.  
 DE PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DE Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 DE Zinc-finger; Multigene family; Alternative splicing; 3D-structure.  
 DE MODULATING (BY SIMILARITY).  
 FT DOMAIN 1 204  
 FT NUCLEAR RECEPTOR-TYPE.  
 FT ZN BIND 205 270  
 FT ZN FING 205 225  
 FT ZN FING 241 265  
 FT HINGE 271 330  
 FT DOMAIN 331 533  
 FT LIGAND-BINDING (BY SIMILARITY).  
 FT DOMAIN 61 198  
 FT PRO-RICH.  
 FT CONFLICT 112 112  
 FT S -> T (IN REF. 2).  
 SQ SEQUENCE 533 AA; 56921 MW; D0069FE93AC16A04 CRC64;  
 Query Match 14.6%; Score 150.5; DB 1; Length 533;  
 Best Local Similarity 31.0%; Pred. No. 0.17;  
 Matches 40; Conservative 11; Mismatches 45; Indels 33; Gaps 5;  
 QY 31 NDKPLPQO--OPFAPAMQDNSSONTLQTPPIAPA-----PKPAAPRPLDRESP 81  
 DB 86 SSFNLPQGVPPSPFG-----PPLPSTAPSLGSGGAPPPPMPPPLGSPFP 134  
 QY 82 GVKNLIPVSGSPASSTPLPDPGTGPNSTNNRATVPVSGSNSSSADPKAP----- 133  
 DB 135 VISS-----SMGSGPLPAPPAPFGSGPSSPINSVSLPGSGGPPEDVKPVLGVRLH 190  
 QY 134 -PPPVSSG 141  
 DB 191 CFPFGPGP 199  
 RESULT 10  
 MUC2 HUMAN  
 ID MUC2 HUMAN STANDARD; PRT; 5179 AA.  
 AC Q02817; Q14878;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Mucin 2 precursor (Intestinal mucin 2).  
 GN MUC2 OR SMUC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RA TISSUE=Intestine;  
 RA MEDLINE=94132002; PubMed=8300571;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Siddiki B., Kim Y.S.;  
 RT "Molecular cloning of human intestinal mucin (MUC2) cDNA.  
 RT Identification of the amino terminus and overall sequence similarity  
 RT to prepro-von Willebrand factor";  
 RL J. Biol. Chem. 269:2440-2446 (1994).  
 RL [2]  
 RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.  
 RP TISSUE=Colon;  
 RX MEDLINE=93016075; PubMed=1400449;  
 RA Gum J.R. Jr., Hicks J.W., Toribara N.W., Roche E.-M., Lagace R.E.,  
 RA Kim Y.S.;  
 RT "The human MUC2 intestinal mucin has cysteine-rich subdomains located  
 RT both upstream and downstream of its central repetitive region.";  
 RL J. Biol. Chem. 267:21375-21383 (1992).  
 RN [3]  
 RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.

RX MEDLINE=91359717; PubMed=1885763;  
 RA Toribara N.W., Gum J.R. Jr., Culhane P.J., Lagace R.E., Hicks J.W.,  
 RA Petersen G.M., Kim Y.S.;  
 RA "MUC-2 human small intestinal mucin gene structure. Repeated arrays  
 and polymorphism."  
 RL J. Clin. Invest. 88:1005-1013(1991).  
 CC !- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND  
 CC OTHER MUCUS MEMBRANE-CONTAINING ORGANS. THOUGHT TO PROVIDE A  
 CC PROTECTIVE, LUBRICATING BARRIER AGAINST PARTICLES AND INFECTIOUS  
 CC AGENTS AT MUCOSAL SURFACES.  
 CC !- SUBUNIT: MULTIMERIC.  
 CC !- SUBCELLULAR LOCATION: Secreted.  
 CC !- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,  
 CC BRONCHUS, CERVIX AND GALL BLADDER  
 CC !- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR  
 CC INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).  
 CC !- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
 CC VARIES AMONG DIFFERENT ALLELES.  
 CC !- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.  
 CC !- SIMILARITY: Contains 1 TIL (trypsin inhibitory-like) domain.  
 CC !- SIMILARITY: Contains 2 WFCC domains.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L21998; AAB95295.1; -  
 CC EMBL; M74027; AAA59875.1; -  
 CC EMBL; M94131; AAA59163.1; -  
 CC EMBL; M94132; AAA59164.1; -  
 CC PIR; A49963; A43932.  
 CC Genew; HGNC:7512; MUC2.  
 CC MIM; 158370; -  
 CC GO; GO:0005803; C:secretory vesicle, TAS.  
 CC InterPro; IPR006208; Cys\_knot.  
 CC InterPro; IPR006207; Cys\_knot\_C.  
 CC InterPro; IPR006209; EGF\_like.  
 CC InterPro; IPR002819; TIL\_Cyrich.  
 CC InterPro; IPR001007; VWF\_C.  
 CC InterPro; IPR001846; VWF\_D.  
 CC Pfam; PF00007; Cys\_Knot; 1.  
 CC Pfam; PF01826; TIL; 1.  
 CC Pfam; PF00093; vwc; 1.  
 CC Pfam; PF00094; vwd; 4.  
 CC SMART; SM00214; VWC; 2.  
 CC SMART; SM00216; VWD; 4.  
 CC PROSITE; PS01185; CTCK\_1; 1.  
 CC PROSITE; PS01225; CTCK\_2; 1.  
 CC PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 CC PROSITE; PS01208; VWF\_1; 2.  
 CC PROSITE; PS0184; VWF\_2; 2.  
 CC KW Glycoprotein; Repeat; Signal.  
 CC SIGNAL 1 20  
 CC CHAIN 21 5179  
 CC DOMAIN 1401 1747  
 CC REPEAT 1401 1416  
 CC REPEAT 1417 1432  
 CC REPEAT 1433 1448  
 CC REPEAT 1449 1464  
 CC REPEAT 1465 1471  
 CC REPEAT 1472 1478  
 CC REPEAT 1479 1494  
 CC REPEAT 1495 1517  
 CC REPEAT 1518 1533  
 CC REPEAT 1534 1556  
 CC REPEAT 1557 1572  
 CC REPEAT 1573 1596  
 CC REPEAT 1597 1612  
 CC REPEAT 1613 1635

REPEAT 1636 1651  
 REPEAT 1652 1675  
 REPEAT 1676 1683  
 REPEAT 1684 1699  
 REPEAT 1700 1715  
 REPEAT 1716 1731  
 REPEAT 1732 1747  
 DOMAIN 1748 1763  
 DOMAIN 1764 1779  
 DOMAIN 1780 1795  
 DISULFID 1796 1811  
 DISULFID 1812 1827  
 DISULFID 1828 1843  
 DISULFID 1844 1859  
 CARBOHYD 1860 1875  
 CARBOHYD 1876 1891  
 CARBOHYD 1892 1907  
 CARBOHYD 1908 1923  
 CARBOHYD 1924 1939  
 CARBOHYD 1940 1955  
 CARBOHYD 1956 1971  
 CARBOHYD 1972 1987  
 CARBOHYD 1988 2003  
 CARBOHYD 2004 2019  
 CARBOHYD 2020 2035  
 CARBOHYD 2036 2051  
 CARBOHYD 2052 2067  
 CARBOHYD 2068 2083  
 CARBOHYD 2084 2099  
 CARBOHYD 2100 2115  
 CARBOHYD 2116 2131  
 CARBOHYD 2132 2147  
 CARBOHYD 2148 2163  
 CARBOHYD 2164 2179  
 CARBOHYD 2180 2195  
 CARBOHYD 2196 2211  
 CARBOHYD 2212 2227  
 CARBOHYD 2228 2243  
 CARBOHYD 2244 2259  
 CARBOHYD 2260 2275  
 CARBOHYD 2276 2291  
 CARBOHYD 2292 2307  
 CARBOHYD 2308 2323  
 CARBOHYD 2324 2339  
 CARBOHYD 2340 2355  
 CARBOHYD 2356 2371  
 CARBOHYD 2372 2387  
 CARBOHYD 2388 2403  
 CARBOHYD 2404 2419  
 CARBOHYD 2420 2435  
 CARBOHYD 2436 2451  
 CARBOHYD 2452 2467  
 CARBOHYD 2468 2483  
 CARBOHYD 2484 2499  
 CARBOHYD 2500 2515  
 CARBOHYD 2516 2531  
 CARBOHYD 2532 2547  
 CARBOHYD 2548 2563  
 CARBOHYD 2564 2579  
 CARBOHYD 2580 2595  
 CARBOHYD 2596 2611  
 CARBOHYD 2612 2627  
 CARBOHYD 2628 2643  
 CARBOHYD 2644 2659  
 CARBOHYD 2660 2675  
 CARBOHYD 2676 2691  
 CARBOHYD 2692 2707  
 CARBOHYD 2708 2723  
 CARBOHYD 2724 2739  
 CARBOHYD 2740 2755  
 CARBOHYD 2756 2771  
 CARBOHYD 2772 2787  
 CARBOHYD 2788 2803  
 CARBOHYD 2804 2819  
 CARBOHYD 2820 2835  
 CARBOHYD 2836 2851  
 CARBOHYD 2852 2867  
 CARBOHYD 2868 2883  
 CARBOHYD 2884 2899  
 CARBOHYD 2900 2915  
 CARBOHYD 2916 2931  
 CARBOHYD 2932 2947  
 CARBOHYD 2948 2963  
 CARBOHYD 2964 2979  
 CARBOHYD 2980 2995  
 CARBOHYD 2996 3011  
 CARBOHYD 3012 3027  
 CARBOHYD 3028 3043  
 CARBOHYD 3044 3059  
 CARBOHYD 3060 3075  
 CARBOHYD 3076 3091  
 CARBOHYD 3092 3107  
 CARBOHYD 3108 3123  
 CARBOHYD 3124 3139  
 CARBOHYD 3140 3155  
 CARBOHYD 3156 3171  
 CARBOHYD 3172 3187  
 CARBOHYD 3188 3203  
 CARBOHYD 3204 3219  
 CARBOHYD 3220 3235  
 CARBOHYD 3236 3251  
 CARBOHYD 3252 3267  
 CARBOHYD 3268 3283  
 CARBOHYD 3284 3299  
 CARBOHYD 3300 3315  
 CARBOHYD 3316 3331  
 CARBOHYD 3332 3347  
 CARBOHYD 3348 3363  
 CARBOHYD 3364 3379  
 CARBOHYD 3380 3395  
 CARBOHYD 3396 3411  
 CARBOHYD 3412 3427  
 CARBOHYD 3428 3443  
 CARBOHYD 3444 3459  
 CARBOHYD 3460 3475  
 CARBOHYD 3476 3491  
 CARBOHYD 3492 3507  
 CARBOHYD 3508 3523  
 CARBOHYD 3524 3539  
 CARBOHYD 3540 3555  
 CARBOHYD 3556 3571  
 CARBOHYD 3572 3587  
 CARBOHYD 3588 3603  
 CARBOHYD 3604 3619  
 CARBOHYD 3620 3635  
 CARBOHYD 3636 3651  
 CARBOHYD 3652 3667  
 CARBOHYD 3668 3683  
 CARBOHYD 3684 3699  
 CARBOHYD 3700 3715  
 CARBOHYD 3716 3731  
 CARBOHYD 3732 3747  
 CARBOHYD 3748 3763  
 CARBOHYD 3764 3779  
 CARBOHYD 3780 3795  
 CARBOHYD 3796 3811  
 CARBOHYD 3812 3827  
 CARBOHYD 3828 3843  
 CARBOHYD 3844 3859  
 CARBOHYD 3860 3875  
 CARBOHYD 3876 3891  
 CARBOHYD 3892 3907  
 CARBOHYD 3908 3923  
 CARBOHYD 3924 3939  
 CARBOHYD 3940 3955  
 CARBOHYD 3956 3971  
 CARBOHYD 3972 3987  
 CARBOHYD 3988 4003  
 CARBOHYD 4004 4019  
 CARBOHYD 4020 4035  
 CARBOHYD 4036 4051  
 CARBOHYD 4052 4067  
 CARBOHYD 4068 4083  
 CARBOHYD 4084 4099  
 CARBOHYD 4100 4115  
 CARBOHYD 4116 4131  
 CARBOHYD 4132 4147  
 CARBOHYD 4148 4163  
 CARBOHYD 4164 4179  
 CARBOHYD 4180 4195  
 CARBOHYD 4196 4211  
 CARBOHYD 4212 4227  
 CARBOHYD 4228 4243  
 CARBOHYD 4244 4259  
 CARBOHYD 4260 4275  
 CARBOHYD 4276 4291  
 CARBOHYD 4292 4307  
 CARBOHYD 4308 4323  
 CARBOHYD 4324 4339  
 CARBOHYD 4340 4355  
 CARBOHYD 4356 4371  
 CARBOHYD 4372 4387  
 CARBOHYD 4388 4403  
 CARBOHYD 4404 4419  
 CARBOHYD 4420 4435  
 CARBOHYD 4436 4451  
 CARBOHYD 4452 4467  
 CARBOHYD 4468 4483  
 CARBOHYD 4484 4499  
 CARBOHYD 4500 4515  
 CARBOHYD 4516 4531  
 CARBOHYD 4532 4547  
 CARBOHYD 4548 4563  
 CARBOHYD 4564 4579  
 CARBOHYD 4580 4595  
 CARBOHYD 4596 4611  
 CARBOHYD 4612 4627  
 CARBOHYD 4628 4643  
 CARBOHYD 4644 4659  
 CARBOHYD 4660 4675  
 CARBOHYD 4676 4691  
 CARBOHYD 4692 4707  
 CARBOHYD 4708 4723  
 CARBOHYD 4724 4739  
 CARBOHYD 4740 4755  
 CARBOHYD 4756 4771  
 CARBOHYD 4772 4787  
 CARBOHYD 4788 4803  
 CARBOHYD 4804 4819  
 CARBOHYD 4820 4835  
 CARBOHYD 4836 4851  
 CARBOHYD 4852 4867  
 CARBOHYD 4868 4883  
 CARBOHYD 4884 4899  
 CARBOHYD 4900 4915  
 CARBOHYD 4916 4931  
 CARBOHYD 4932 4947  
 CARBOHYD 4948 4963  
 CARBOHYD 4964 4979  
 CARBOHYD 4980 4995  
 CARBOHYD 4996 5011  
 CARBOHYD 5012 5027  
 CARBOHYD 5028 5043  
 CARBOHYD 5044 5059  
 CARBOHYD 5060 5075  
 CARBOHYD 5076 5091  
 CARBOHYD 5092 5107  
 CARBOHYD 5108 5123  
 CARBOHYD 5124 5139  
 CARBOHYD 5140 5155  
 CARBOHYD 5156 5171  
 CARBOHYD 5172 5187  
 CARBOHYD 5188 5203  
 CARBOHYD 5204 5219  
 CARBOHYD 5220 5235  
 CARBOHYD 5236 5251  
 CARBOHYD 5252 5267  
 CARBOHYD 5268 5283  
 CARBOHYD 5284 5299  
 CARBOHYD 5300 5315  
 CARBOHYD 5316 5331  
 CARBOHYD 5332 5347  
 CARBOHYD 5348 5363  
 CARBOHYD 5364 5379  
 CARBOHYD 5380 5395  
 CARBOHYD 5396 5411  
 CARBOHYD 5412 5427  
 CARBOHYD 5428 5443  
 CARBOHYD 5444 5459  
 CARBOHYD 5460 5475  
 CARBOHYD 5476 5491  
 CARBOHYD 5492 5507  
 CARBOHYD 5508 5523  
 CARBOHYD 5524 5539  
 CARBOHYD 5540 5555  
 CARBOHYD 5556 5571  
 CARBOHYD 5572 5587  
 CARBOHYD 5588 5603  
 CARBOHYD 5604 5619  
 CARBOHYD 5620 5635  
 CARBOHYD 5636 5651  
 CARBOHYD 5652 5667  
 CARBOHYD 5668 5683  
 CARBOHYD 5684 5699  
 CARBOHYD 5700 5715  
 CARBOHYD 5716 5731  
 CARBOHYD 5732 5747  
 CARBOHYD 5748 5763  
 CARBOHYD 5764 5779  
 CARBOHYD 5780 5795  
 CARBOHYD 5796 5811  
 CARBOHYD 5812 5827  
 CARBOHYD 5828 5843  
 CARBOHYD 5844 5859  
 CARBOHYD 5860 5875  
 CARBOHYD 5876 5891  
 CARBOHYD 5892 5907  
 CARBOHYD 5908 5923  
 CARBOHYD 5924 5939  
 CARBOHYD 5940 5955  
 CARBOHYD 5956 5971  
 CARBOHYD 5972 5987  
 CARBOHYD 5988 6003  
 CARBOHYD 6004 6019  
 CARBOHYD 6020 6035  
 CARBOHYD 6036 6051  
 CARBOHYD 6052 6067  
 CARBOHYD 6068 6083  
 CARBOHYD 6084 6099  
 CARBOHYD 6100 6115  
 CARBOHYD 6116 6131  
 CARBOHYD 6132 6147  
 CARBOHYD 6148 6163  
 CARBOHYD 6164 6179  
 CARBOHYD 6180 6195  
 CARBOHYD 6196 6211  
 CARBOHYD 6212 6227  
 CARBOHYD 6228 6243  
 CARBOHYD 6244 6259  
 CARBOHYD 6260 6275  
 CARBOHYD 6276 6291  
 CARBOHYD 6292 6307  
 CARBOHYD 6308 6323  
 CARBOHYD 6324 6339  
 CARBOHYD 6340 6355  
 CARBOHYD 6356 6371  
 CARBOHYD 6372 6387  
 CARBOHYD 6388 6403  
 CARBOHYD 6404 6419  
 CARBOHYD 6420 6435  
 CARBOHYD 6436 6451  
 CARBOHYD 6452 6467  
 CARBOHYD 6468 6483  
 CARBOHYD 6484 6499  
 CARBOHYD 6500 6515  
 CARBOHYD 6516 6531  
 CARBOHYD 6532 6547  
 CARBOHYD 6548 6563  
 CARBOHYD 6564 6579  
 CARBOHYD 6580 6595  
 CARBOHYD 6596 6611  
 CARBOHYD 6612 6627  
 CARBOHYD 6628 6643  
 CARBOHYD 6644 6659  
 CARBOHYD 6660 6675  
 CARBOHYD 6676 6691  
 CARBOHYD 6692 6707  
 CARBOHYD 6708 6723  
 CARBOHYD 6724 6739  
 CARBOHYD 6740 6755  
 CARBOHYD 6756 6771  
 CARBOHYD 6772 6787  
 CARBOHYD 6788 6803  
 CARBOHYD 6804 6819  
 CARBOHYD 6820 6835  
 CARBOHYD 6836 6851  
 CARBOHYD 6852 6867  
 CARBOHYD 6868 6883  
 CARBOHYD 6884 6899  
 CARBOHYD 6900 6915  
 CARBOHYD 6916 6931  
 CARBOHYD 6932 6947  
 CARBOHYD 6948 6963  
 CARBOHYD 6964 6979  
 CARBOHYD 6980 6995  
 CARBOHYD 6996 7011  
 CARBOHYD 7012 7027  
 CARBOHYD 7028 7043  
 CARBOHYD 7044 7059  
 CARBOHYD 7060 7075  
 CARBOHYD 7076 7091  
 CARBOHYD 7092 7107  
 CARBOHYD 7108 7123  
 CARBOHYD 7124 7139  
 CARBOHYD 7140 7155  
 CARBOHYD 7156 7171  
 CARBOHYD 7172 7187  
 CARBOHYD 7188 7203  
 CARBOHYD 7204 7219  
 CARBOHYD 7220 7235  
 CARBOHYD 7236 7251  
 CARBOHYD 7252 7267  
 CARBOHYD 7268 7283  
 CARBOHYD 7284 7299  
 CARBOHYD 7300 7315  
 CARBOHYD 7316 7331  
 CARBOHYD 7332 7347  
 CARBOHYD 7348 7363  
 CARBOHYD 7364 7379  
 CARBOHYD 7380 7395  
 CARBOHYD 7396 7411  
 CARBOHYD 7412 7427  
 CARBOHYD 7428 7443  
 CARBOHYD 7444 7459  
 CARBOHYD 7460 7475  
 CARBOHYD 7476 7491  
 CARBOHYD 7492 7507  
 CARBOHYD 7508 7523  
 CARBOHYD 7524 7539  
 CARBOHYD 7540 7555  
 CARBOHYD 7556 7571  
 CARBOHYD 7572 7587  
 CARBOHYD 7588 7603  
 CARBOHYD 7604 7619  
 CARBOHYD 7620 7635  
 CARBOHYD 7636 7651  
 CARBOHYD 7652 7667  
 CARBOHYD 7668 7683  
 CARBOHYD 7684 7699  
 CARBOHYD 7700 7715  
 CARBOHYD 7716 7731  
 CARBOHYD 7732 7747  
 CARBOHYD 7748 7763  
 CARBOHYD 7764 7779  
 CARBOHYD 7780 7795  
 CARBOHYD 7796 7811  
 CARBOHYD 7812 7827  
 CARBOHYD 7828 7843  
 CARBOHYD 7844 7859  
 CARBOHYD 7860 7875  
 CARBOHYD 7876 7891  
 CARBOHYD 7892 7907  
 CARBOHYD 7908 7923  
 CARBOHYD 7924 7939  
 CARBOHYD 7940 7955  
 CARBOHYD 7956 7971  
 CARBOHYD 7972 7987  
 CARBOHYD 7988 8003  
 CARBOHYD 8004 8019  
 CARBOHYD 8020 8035  
 CARBOHYD 8036 8051  
 CARBOHYD 8052 8067  
 CARBOHYD 8068 8083  
 CARBOHYD 8084 8099  
 CARBOHYD 8100 8115  
 CARBOHYD 8116 8131  
 CARBOHYD 8132 8147  
 CARBOHYD 8148 8163  
 CARBOHYD 8164 8179  
 CARBOHYD 8180 8195  
 CARBOHYD 8196 8211  
 CARBOHYD 8212 8227  
 CARBOHYD 8228 8243  
 CARBOHYD 8244 8259  
 CARBOHYD 8260 8275  
 CARBOHYD 8276 8291  
 CARBOHYD 8292 8307  
 CARBOHYD 8308 8323  
 CARBOHYD 8324 8339  
 CARBOHYD 8340 8355  
 CARBOHYD 8356 8371  
 CARBOHYD 8372 8387  
 CARBOHYD 8388 8403  
 CARBOHYD 8404 8419  
 CARBOHYD 8420 8435  
 CARBOHYD 8436 8451  
 CARBOHYD 8452 8467  
 CARBOHYD 8468 8483  
 CARBOHYD 8484 8499  
 CARBOHYD 8500 8515  
 CARBOHYD 8516 8531  
 CARBOHYD 8532 8547  
 CARBOHYD 8548 8563  
 CARBOHYD 8564 8579  
 CARBOHYD 8580 8595  
 CARBOHYD 8596 8611  
 CARBOHYD 8612 8627  
 CARBOHYD 8628 8643  
 CARBOHYD 8644 8659  
 CARBOHYD 8660 8675  
 CARBOHYD 8676 8691  
 CARBOHYD 8692 8707  
 CARBOHYD 8708 8723  
 CARBOHYD 8724 8739  
 CARBOHYD 8740 8755  
 CARBOHYD 8756 8771  
 CARBOHYD 8772 8787  
 CARBOHYD 8788 8803  
 CARBOHYD 8804 8819  
 CARBOHYD 8820 8835  
 CARBOHYD 8836 8851  
 CARBOHYD 8852 8867  
 CARBOHYD 8868 8883  
 CARBOHYD 8884 8899  
 CARBOHYD 8900 8915  
 CARBOHYD 8916 8931  
 CARBOHYD 8932 8947  
 CARBOHYD 8948 8963  
 CARBOHYD 8964 8979  
 CARBOHYD 8980 8995  
 CARBOHYD 8996 9011  
 CARBOHYD 9012 9027  
 CARBOHYD 9028 9043  
 CARBOHYD 9044 9059  
 CARBOHYD 9060 9075  
 CARBOHYD 9076 9091  
 CARBOHYD 9092 9107  
 CARBOHYD 9108 9123  
 CARBOHYD 9124 9139  
 CARBOHYD 9140 9155  
 CARBOHYD 9156 9171  
 CARBOHYD 9172 9187  
 CARBOHYD 9188 9203  
 CARBOHYD 9204 9219  
 CARBOHYD 9220 9235  
 CARBOHYD 9236 9251  
 CARBOHYD 9252 9267  
 CARBOHYD 9268 9283  
 CARBOHYD 9284 9299  
 CARBOHYD 9300 9315  
 CARBOHYD 9316 9331  
 CARBOHYD 9332 9347  
 CARBOHYD 9348 9363  
 CARBOHYD 9364 9379  
 CARBOHYD 9380 9395  
 CARBOHYD 9396 9411  
 CARBOHYD 9412 9427  
 CARBOHYD 9428 9443  
 CARBOHYD 9444 9459  
 CARBOHYD 9460 9475  
 CARBOHYD 9476 9491  
 CARBOHYD 9492 9507  
 CARBOHYD 9508 9523  
 CARBOHYD 9524 9539  
 CARBOHYD 9540 9555  
 CARBOHYD 9556 9571  
 CARBOHYD 9572 9587  
 CARBOHYD 9588 9603  
 CARBOHYD 9604 9619  
 CARBOHYD 9620 9635  
 CARBOHYD 9636 9651  
 CARBOHYD 9652 9667  
 CARBOHYD 9668 9683  
 CARBOHYD 9684 9699  
 CARBOHYD 9700 9715  
 CARBOHYD 9716 9731  
 CARBOHYD 9732 9747  
 CARBOHYD 9748 9763  
 CARBOHYD 9764 9779  
 CARBOHYD 9780 9795  
 CARBOHYD 9796 9811  
 CARBOHYD 9812 9827  
 CARBOHYD 9828 9843  
 CARBOHYD 9844 9859  
 CARBOHYD 9860 9875  
 CARBOHYD 9876 9891  
 CARBOHYD 9892 9907  
 CARBOHYD 9908 9923  
 CARBOHYD 9924 9939  
 CARBOHYD 9940 9955  
 CARBOHYD 9956 9971  
 CARBOHYD 9972 9987  
 CARBOHYD 9988 1003  
 CARBOHYD 1004 1019  
 CARBOHYD 1020 1035  
 CARBOHYD 1036 1051  
 CARBOHYD 1052 1067  
 CARBOHYD 1068 1083  
 CARBOHYD 1084 1099  
 CARBOHYD 1100 1115  
 CARBOHYD 1116 1131  
 CARBOHYD 1132 1147  
 CARBOHYD 1148 1163  
 CARBOHYD 1164 1179  
 CARBOHYD 1180 1195  
 CARBOHYD 1196 1211  
 CARBOHYD 1212 1227  
 CARBOHYD 1228 1243  
 CARBOHYD 1244 1259  
 CARBOHYD 1260 1275  
 CARBOHYD 1276 1291  
 CARBOHYD 1292 1307  
 CARBOHYD 1308 1323  
 CARBOHYD 1324 1339  
 CARBOHYD 1340 1355  
 CARBOHYD 1356 1371  
 CARBOHYD 1372 1387  
 CARBOHYD 1388 1403  
 CARBOHYD 1404 1419  
 CARBOHYD 1420 1435  
 CARBOHYD 1436 1451  
 CARBOHYD 1452 1467  
 CARBOHYD 1468 1483  
 CARBOHYD 1484 1499  
 CARBOHYD 1500 1515  
 CARBOHYD 1516 1531  
 CARBOHYD 1532 1547  
 CARBOHYD 1548 1563  
 CARBOHYD 1564 1579  
 CARBOHYD 1580 1595  
 CARBOHYD 1596 1611  
 CARBOHYD 1612 1627  
 CARBOHYD 1628 1643  
 CARBOHYD 1644 1659  
 CARBOHYD 1660 1675  
 CARBOHYD 1676 1691  
 CARBOHYD 1692 1707  
 CARBOHYD 1708 1723  
 CARBOHYD 1724 1739  
 CARBOHYD 1740 1755  
 CARBOHYD 1756 1771  
 CARBOHYD 1772 1787  
 CARBOHYD 1788 1803  
 CARBOHYD 1804 1819  
 CARBOHYD 1820 1835  
 CARBOHYD 1836 1851  
 CARBOHYD 1852 1867  
 CARBOHYD 1868 1883  
 CARBOHYD 1884 1899  
 CARBOHYD 1900 1915  
 CARBOHYD 1916 1931  
 CARBOHYD 1932 1947  
 CARBOHYD 1948 1963  
 CARBOHYD 1964 1979  
 CARBOHYD 1980 1995  
 CARBOHYD 1996 2011  
 CARBOHYD 2012 2027  
 CARBOHYD 2028 2043  
 CARBOHYD 2044 2059  
 CARBOHYD 2060 2075  
 CARBOHYD 2076 2091  
 CARBOHYD 2092 2107  
 CARBOHYD 2108 2123  
 CARBOHYD 2124 2139  
 CARBOHY

DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-1999 (Rel. 38, Last annotation update)  
 DE Pistil-specific extensin-like protein precursor (PELP).  
 OS Nicotiana tabacum (Common tobacco).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 CC Asteridae; Lamiales; Solanales; Solanaceae; Nicotiana.  
 CX NCBI\_TaxID=4097;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=cv. Petit Havana; TISSUE=Pistil;  
 RA MEDLINE=93005740; PubMed=192607;  
 RX Goldman M.H., Pezotti M., Seurinck J., Mariani C.;  
 RT "Developmental expression of tobacco pistil-specific genes encoding  
 RT novel extensin-like proteins.";  
 RL Plant Cell 4:1041-1051(1992).  
 CC -!- TISSUE SPECIFICITY: PISTIL (STIGMA AND STYLE TISSUE).  
 CC -!- DEVELOPMENTAL STAGE: EXPRESSION BEGINS IN FLORAL BUDS AFTER  
 CC PISTIL DIFFERENTIATION AND LEVELS GRADUALLY INCREASE DURING  
 CC FLOWER DEVELOPMENT TOWARD ANTHESIS. LEVELS GRADUALLY DECREASE  
 CC AFTER POLLINATION AND ARE ABSENT BY THE SIXTH DAY AFTER  
 CC POLLINATION.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z14019; CAA78397.1;  
 DR PIR; JQ1696; JQ1696.  
 DR InterPro; IPR006041; Ole\_el\_ext.  
 DR InterPro; IPR002965; P-rich\_extensin.  
 DR InterPro; IPR003488; Pistil\_extensin.  
 DR Pfam; PF01190; Pollen\_Ole\_e\_I; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PRINTS; PR01218; PSTLEXTENSIN.  
 KW Structural protein; Signal; Repeat; Glycoprotein.  
 FT SIGNAL 1 23  
 FT CHAIN 24 426 PISTIL-SPECIFIC EXTENSIN-LIKE PROTEIN.  
 FT DOMAIN 69 182 4 X 5 AA REPEATS OF S-P(4).  
 FT REPEAT 69 73 1  
 FT REPEAT 76 80 2  
 FT REPEAT 83 87 3  
 FT REPEAT 178 182 4  
 FT CARBOHYD 310 310 N-LINKED (GLCNAC...) (POTENTIAL).  
 SQ SEQUENCE 426 AA; 41278 MW; 51A495CC94017812 CRC64;  
 Query Match 14.5%; Score 149; DB 1; Length 426;  
 Best Local Similarity 29.9%; Pred. No. 0.17;  
 Matches 40; Conservative 14; Mismatches 54; Indels 26; Gaps 5;  
 QY 31 NDPKLPQPPAPANQDQNS--SQNTRLOPTTPIAPAPKPAAPRLDPSGIVENKLI 88  
 DB 140 SSGPLVKPPPPPSFCPSPPQSQAKQPQPP-PAKQSPPPPPPPVVKAPSP----- 191  
 QY 89 PSVGSASSTTLPDGT--GNSTPMNRAVTPVSGSSSSADPKAPPPPPVVS----- 139  
 DB 192 ----SPAKPPPPPPVVKAPSPSPATQPPKPPPPPPRAKPSLLPPPPVAVPPWTPS 247  
 QY 140 ---SCEPPTLGENP 150  
 DB 248 PSFAEPPIIAPFP 261  
 RESULT 12  
 CSP\_PLABA  
 ID CSP\_PLABA STANDARD; PRT; 347 AA.  
 AC P23033;  
 DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Circumsporozoite protein precursor (CS).  
 OS Plasmodium berghei (strain Anka).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxID=5823;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=90221834; PubMed=2183186;  
 RA Lockyer M.J., Davies C.S., Suhrbier A., Sinden R.E.;  
 RT "Nucleotide sequence of the plasmodium berghei circumsporozoite  
 RT protein gene from the ANKA clone 2.34L.";  
 RL Nucleic Acids Res. 18:376-376(1990).  
 CC -!- FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOOROZOITE (THE INFECTIVE STAGE OF THE  
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -!- MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X17606; CAA35608.1;  
 DR PIR; S07873; OZZQBK.  
 DR InterPro; IPR003067; Crcmsporozoite.  
 DR InterPro; IPR000884; TSP1.  
 DR Pfam; PF00090; TSP\_1; 1.  
 DR PRINTS; PR01103; ECHMSPRZOITE.  
 DR SMART; SM00209; TSP1; 1.  
 DR PROSITE; PS50092; TSP1; 1.  
 KW Malaria; Sporozoite; Repeat; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 347 CIRCUMSPOROZOITE PROTEIN.  
 FT DOMAIN 93 204 13 X 8 AA REPEATS.  
 FT DOMAIN 214 247 17 X 2 AA REPEATS OF P-Q.  
 FT DOMAIN 274 325 TSP TYPE-1.  
 SQ SEQUENCE 347 AA; 37776 MW; 0EC240EE35681AF8 CRC64;  
 Query Match 14.3%; Score 147; DB 1; Length 347;  
 Best Local Similarity 29.9%; Pred. No. 0.18;  
 Matches 47; Conservative 10; Mismatches 68; Indels 32; Gaps 8;  
 QY 8 QVNNKNTSTAPLNTQISALRNDPKLPQPPAPANQDQNSQNTRLQPTTP-----IP 63  
 DB 76 KVEKNKIER-----NNKLA--QPPPPNPNNDPPPNPNND-----PPPNPNNDPP 118  
 QY 64 APAPKPAAPRLDPSGIVENKLI PSVGSPPA--SSTPLPDGTGP-----NSTPNRA 115  
 DB 119 PPNNDPPPNNDPPPNNDPPPNNDPPPNNDPPPNNDPPPNNDPPPNNDPPPNNA 178  
 QY 116 VTFVSGSSSSADPKA--PPPPVSVSGEPTLGENP 150  
 DB 179 NDPPPPNPN- DPAPENANDPPPNPNNDPPPNNDPPPNNDPPPNNDPPPNNA 214  
 RESULT 13  
 AFG\_ARATH  
 ID AFG\_ARATH STANDARD; PRT; 534 AA.  
 AC P40602; Q93214; Q9LNT8;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Anter-specific proline-rich protein APG precursor.  
 GN APG OR ATIG20130 OR T20H2.9  
 OS Arabidopsis thaliana (Mouse-ear cress).



```

FT ACT_SITE 511 511 POTENTIAL.
FT CONFLICT 77 77 S -> P (IN RFP. 1).
FT CONFLICT 141 141 E -> A (IN RFP. 1).
FT CONFLICT 325 325 E -> H (IN RFP. 1).
SQ SEQUENCE 534 AA; 58007 MW; BA851DC3CF7429DB CRC64;

Query Match
Best Local Similarity 34.2%; Pred. NO. 0.29;
Matches 39; Conservative 10; Mismatches 45; Indels 20; Gaps 5;

QY 33 PKPLPQQPAPPA-NQDQSSNQTRLQPTTP-----IPAPAPKPAAPRPRLDRESGVEN 85
DB 93 FKTQPKPPFASDSSPCSPSPKQPKQKVPFPPACFPPTPKQPKQKAPPPEPKAPPKPK 152
QY 86 KLIPSGVGSASST--PLPDDCTGPNSTNNRAVTPVQSGSSSSADPKAPPPPP 137
DB 153 VCFPSPPKPPAPTEKVPFPHGPKKAP--AATP-----APSPKPAFSP 195

RESULT 14
SHK1_RAT
ID SHK1_RAT STANDARD; PRT; 2167 AA.
AC Q9W4G8; Q9QZ28; Q9WU13; Q9WUE9;
DT 28-FEB-2003 (Rel. 41, Created);
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
DE interacting protein) (SPANK-1) (Synomem) (Isomactatcin receptor
DE interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI TaxID=10116;
[1]
RP RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
RP DLG4.
RC TISSUE=Brain;
RA MEDLINE=99419021; PubMed=10488079;
RX Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "synomem, a novel neuronal protein interacting with synapse-associated
RT protein 90/postsynaptic density-95-associated protein.";
RN J. Biol. Chem. 274:27463-27466(1999).
[2]
RP RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=9930650; PubMed=10433268;
RA Naibitt S., Kim E., Tu J.C., Xiao B., Sala C., Valschanoff J.,
RX Weinberg R.J., Worley P.F., Sheng M.;
RT "shank, a novel family of postsynaptic density proteins that binds to
RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RN Neuron 23:569-582(1999).
[3]
RP RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "the G protein-coupled receptor C11 interacts directly with proteins
RT of the shank family.";
RN J. Biol. Chem. 275:36204-36210(2000).
[4]
RP RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
RP DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naibitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RT "Characterization of the Shank family of synaptic proteins. Multiple
RT genes, alternative splicing, and differential expression in brain and
RT development.";
RN J. Biol. Chem. 274:29510-29518(1999).
[5]
RP RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;

```



```

SQ      SEQUENCE      2167 AA;      226333 MW;      3F478B5A7B18ABA86 CRC64;

Query Match
Best local similarity      23.2%;      Score 146;      DB 1;      Length 2167;
Matches      56;      Conservative      35;      Mismatches      82;      Indels      62;      Gaps      11;

Qy      6      HIQINSNKTERSTAPLNTQISALRNDP-----KPLPQQPPAPANQDNSSQNTRLQ      57
      :      :      :      :      :      :      :      :      :      :      :      :
Db      1497      HVRFLEN-----CQARPPFAGTGGSSSTDGQGVPPSPRVLFSTSPRGNGENG-----L      1548

Qy      58      PTEPPIDAP-----KPAAPP-----RPLDRSPGVENKLIPSVGSPASS      97
      :      :      :      :      :      :      :      :      :      :      :      :
Db      1549      PLLVLPPFAPSVDVDGGEFLFAEPPLPPLEFNSPESELPFGPPLPL-PPDPSPATP      1607

Qy      98      TLPP-----DGTGNSPPNNRAVTVSGNSSSSADKAPAPPPVSSGEPPTL      146
      :      :      :      :      :      :      :      :      :      :      :      :
Db      1608      LPNAPPAVAAAPPTLDSTASLTSYDSEVALTQAPAFGDGPPAPFAPGAPAP      1667

Qy      147      GENPD-----GLSQOQLRHRS-----LQTLRIQRM-----LFPDEKFTGAQSGG      189

Db      1668      OCPGPPPPPTDSGI-----EEVDSSRSSDHPLETISSASTLSLSAEGSGMTGVAGG      1720

```

RESULT 15

ID\_CSP\_PLABE ID\_CSP\_STANDARD; PRT: 339 AA.  
 AT P06315;  
 DC 01-JAN-1988 (rel. 06, Created)  
 DT 01-JAN-1988 (rel. 06, Last sequence update)  
 DT 28-FEB-2003 (rel. 41, Last annotation update)  
 DE Circumsporozoite protein precursor (CS).  
 DE Plasmodium berghei  
 DE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5821;  
 RN 11)  
 RP SEQUENCE FROM N.A. PubMed=2432395;  
 RA MEDLINE=87089740; Pubmed=2432395;  
 FX Eichinger D.J.J.; Arnot D.E.; Tas J.P.; Nussenzeig V.; Enea V.;  
 ET "Circumsporozoite Protein of Plasmodium berghei: gene cloning and  
 PT identification of immunodominant epitopes."; Mol  
 PL Mol Cell Biol 6: 37945-37973(1986)  
 CC -! FUNCTION: THE CIRCUMSPOROZOITE PROTEIN IS THE IMMUNODOMINANT  
 CC SURFACE ANTIGEN ON THE SPOOROITE (THE INFECTIVE STAGE OF THE  
 CC MALARIA PARASITE THAT IS TRANSMITTED FROM THE MOSQUITO TO THE  
 CC VERTEBRATE HOST).  
 CC -! MISCELLANEOUS: THE C-TERMINAL REGION IS PROBABLY USED FOR  
 CC ANCHORING THE PROTEIN TO THE CELL MEMBRANE. THE REPEAT SEQUENCES  
 CC WOULD BE THE SURFACE ANTIGEN OF THE ORGANISM.  
 CC -! SIMILARITY: Contains 1 TSP type-1 domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>;  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL, M14135; AAA29577.1; -.  
 CC FIR; A44948; OZZQMB.  
 CC InterPro; IPR003067; Crcmsprzoite.  
 CC InterPro; IPR000884; TSP1.  
 CC Pfam; PF00090; tsp\_1; 1.  
 CC PRINTS; PR01303; CRCMSPRZOITE.  
 CC SMART; SM00209; TSP1; 1.  
 CC PROSITE; PS50092; TSP1; 1.  
 CC Malaria; Sporozoite; Repeat; Signal.  
 CC SIGNAL 1 23 PROBABLE.  
 CC FT CHAIN 24 339 CIRCUMSPOROZOITE PROTEIN.  
 CC FT DOMAIN 93 196 13 X 8 AA TANDEN REPEATS.  
 CC FT DOMAIN 206 238 16 X 2 AA TANDEN REPEATS OF P-Q.  
 CC FT DOMAIN 266 317 TSP TYPE-1.  
 CC SEQUENCE 339 AA; 37138 MW; E806A6D1D9551B CRC64;  
 CC SQ

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 63.4086 Seconds  
(without alignments)  
789.518 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_199\_392

Perfect score: 1028

Sequence: 1 TIVSFHFIQINNNKTERSTA.....LFFDEKFTGSGGQQNP 194

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 23:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protist:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	177.5	17.3	473	10 Q39620	Q39620 chlamydomon
2	168.5	16.4	876	3 Q9P4V9	Q9P4V9 neurospora
3	167.5	16.3	1269	10 Q8W5K6	Q8W5K6 oryza sativ
4	163	15.9	1307	10 Q9LVN1	Q9LVN1 arabidopsis
5	162.5	15.8	874	3 Q36027	Q36027 schizosacch
6	162.5	15.8	1011	3 Q9P444	Q9P444 pneumocysti
7	162.5	15.8	1151	13 Q57580	Q57580 gallus gall
8	161.5	15.7	816	4 Q96G51	Q96G51 homo sapien
9	160.5	15.6	2321	12 Q9DGT6	Q9DGT6 turkey herp
10	160	15.6	488	16 Q8DGT6	Q8DGT6 synchococ
11	160	15.6	839	16 Q9RX57	Q9RX57 deinochococ
12	158.5	15.4	351	10 Q93492	Q93492 chlamydomon
13	158.5	15.4	420	5 Q9VZC2	Q9VZC2 drosophila
14	158.5	15.4	1188	10 Q41805	Q41805 zea mays (m
15	158	15.4	802	11 P70433	P70433 mus musculu
16	157.5	15.3	464	10 Q41645	Q41645 volvox cart

17 156 15.2 1315 10 Q9SPM0  
18 155 15.1 763 2 Q9XDH2  
19 154.5 15.0 386 10 Q9PPQ5  
20 154.5 15.0 731 10 Q65530  
21 153.5 14.9 326 10 Q22514  
22 153.5 14.9 637 4 Q9BSV4  
23 153 14.9 493 11 Q8KI17  
24 152.5 14.8 379 10 Q9SUX2  
25 152 14.8 883 2 Q9R11  
26 152 14.8 956 10 Q9LJ64  
27 151.5 14.7 602 2 Q9AKP3  
28 151.5 14.7 964 5 Q26963  
29 151 14.7 616 4 Q9H6K5  
30 151 14.7 1997 10 Q8LRM7  
31 150.5 14.6 598 16 Q8RKN7  
32 150.5 14.6 699 11 Q8VJ66  
33 150 14.6 309 12 Q8U2B4  
34 150 14.6 409 10 Q9SBM1  
35 150 14.6 556 4 Q15411  
36 150 14.6 1157 4 Q96JK7  
37 150 14.6 3124 4 Q96L91  
38 149 14.5 188 11 Q62106  
39 149 14.5 302 5 Q9GZH1  
40 149 14.5 712 10 Q8RWX5  
41 149 14.5 1201 10 Q9C6S1  
42 148.5 14.4 273 10 Q9ARY7  
43 148.5 14.4 731 10 Q9ZUE0  
44 148.5 14.4 752 10 Q9LQC5  
45 148.5 14.4 1202 4 Q8MXE0

## ALIGNMENTS

### RESULT 1

Q39620 PRELIMINARY; PRT; 473 AA.  
ID AC Q39620;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DE Vsp-3 protein precursor.  
GN VSP-3.  
OS Chlamydomonas reinhardtii.  
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
OC Chlamydomonadaceae; Chlamydomonas.  
OX NCBI\_TaxID=3055;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CC-621;  
RX MEDLINE=95093034; PubMed=8000007;  
RA Woessner J.P., Molendijk A.J., van Egmond P., Klis F.M.,  
EA Goodenough U.W., Haring M.A.,  
FT "Domain conservation in several volvoclean cell wall proteins.";  
RL Plant Mol Biol. 26:947-960(1994).  
DR EMBL; J29029; AAB59531.1; -  
DR InterPro; IPR002965; P1rich\_extensn.  
DR PRINTS; PR01217; PRICHEXTNSN.  
KW Signal.  
FT SIGNAL.  
FT CHAIN 1 31 POTENTIAL.  
SQ SEQUENCE 473 AA; 47532 MW; 38506131FAA674A4 CRC64;

Query Match 17.3%; Score 177.5; DB 10; Length 473;  
Best Local Similarity 30.5%; Pred. No. 1.7e-05;  
Matches 46; Conservative 19; Mismatches 59; Indels 27; Gaps 5;  
QY 14 KTERSTAPLNTQIGALNDFKLPQPPAPANQDSSNTRLOPTPTIPAPAKPAAPP 73  
Db 330 KASPSFSP 376  
QY 74 RPLDRSRQVENKLIPSGVSPASSTPLPPDGTGPNSTPN-NRAVTPVSGGSSSADPKA 132

[illegible]

Query Match	15.88;	Score 162.5;	DB 3;	Length 574;
Best Local Similarity	30.9%;	Pred. No. 0.00021;		
Matches 50;	Conservative 21;	Mismatches 64;	Indels 27;	Gaps 7
QY	7	IQMIGNKTERSTA-----PLNTQISA-----LNDQKLPQQPPAPANQDQ-----NSSQNT 54		
DB	275	IAPVSNVNPAINSTSKPEPLPPSSSVSAALAANKRRPPPPPPSRNRGKPPGIGSGNS 334		
QY	55	RIQFTPP-----IPAPAKPAAPFPDLQKESGVENKLIPIGVGSPA--SSTLPDPDG 104		
DB	335	SLPPPPPPPRGNAAGSIPIPLQOGASAPPPPPPPRSGATGQRPPLSSSRVSNPAPPPA 394		
QY	105	TGPNSTPNNRVATPVSGNSSSADPKAPPPPPVSGEPTL 146		
DB	395	IPGRSAP--ALPLGNAGRST--PVPVTPESLPPGAPSL 431		

RESULT 7	
OS7580	
ID OS7580	PRELIMINARY; PRT; 1151 AA.
AC OS7580;	
DT DT	01-JUN-1998 (TEMPREL.06, Created)
DT DT	01-JUN-1998 (TEMPREL.06, Last sequence update)
DT DT	01-MAR-2003 (TEMPREL.23, Last annotation update)
DE DE	High molecular mass nuclear antigen (Fragment).
DS DS	Gallus gallus (Chicken).
OC OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC OC	Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
CC CC	Gallus.
RN RN	[1] NCBI_TaxID=9031;
RX RX	SEQUENCE FROM N.A.A.
RY RY	MEDLINE=9034404; PubMed=9365273;
RT RT	Shimada K., Harata M., Mizuno S.;
RT RT	"A Nuclear matrix-associated high molecular mass nuclear antigen,



Matches	52;	Conservative	13;	Mismatches	79;	Indels	22;	Gaps	7.
Qy	33	PKP-LPQQPPAPANQNDSNQNTLRQP---	TPIIPAPA---PKPAAP-PRFLDRESPGVE	84					
Dd	222	PTENAPACTATPATACQAIACTTTAQTAA	TAITQTTPATAAAPAQAGGRSAPAPANAQNA	261					
Qy	85	NKLIISVGSFASSTPLPDGTGENSTFNAR	VTVPSYQS--NSSADRPKAAPPVPVSSEGP	143					
Dd	282	GVVFETAVTFESSTFAARSACTTPTTREIN	AQTESASPAPNNSAAAHPNPASEEV-AGRP	340					
Qy	144	PTUGENPGUSQBQLHERSLRSLTLRLDIQR	LMLFPDKBFTCAQSGG	189					
Dd	341	GTAASFESPASPVTVTFRGET-----FD	TASAAGTFSAG	374					
RESULT 12									
Q39492	ID	Q39492	PRELIMINARY; PRT; 351 AA.						
IC	Q39492;								
AD	01-NOV-1996 (tREMBUrel_01, Created)								
DT	01-NOV-1996 (tREMBUrel_01, Last sequence update)								
DD	01-MAR-2003 (tREMBUrel_23, Last annotation update)								
DE	MP6 protein precursor.								
DN	MP6.								
GN	Ciliomonas eugametos.								
OC	Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;								
CC	Chlamydomonadaceae; Chlamydomonas.								
RX	[l]__TaxID=3053;								
ON	NCBI_TaxID=3053;								
CP	SEQUENCE FROM N.A.								
EP	STRAIN-JPEX 10.								
BX	MEDLINE=45030314; PubMed=8000007;								
RA	Woesner J.P., Molendijk A.J.; van Egmond P., Klis F.M.,								
RD	"Domain conservation in several volvoclean cell wall proteins.";								
KL	Plant Moll. Biol. 26:947-960(1994).								
DR	EMBL; L29028; ABAS954.1; ..								
DR	InterPro; IPR002965; P rich extensn.								
DR	PRINTS; PR01217; PRICHEXTENS.								
KW	Signal.								
SIGNAL	1 31 POTENTIAL.								
FF	CHAIN 32 351 POTENTIAL.								
FT	SEQUENCE 351 AA; 35310 MW; 70AD2EFFC74BB68 CRC64;								
SQ									
Query Match 15.4%; Score 158.5; DB 10; Length 351;									
Best Local Similarity 30.9%; Pred. No. 0.00032;									
Matches 46; Conservative 15; Mismatches 71; Indels 17; Gaps 5;									
Qy	9	NISNNKTERTSAPlNQ----	ISALRNDRKLQPPAPANQONS-----SQNRLOQT	59					
Dd	144	SVYNWCIDTRPAPVCNSTFNVTSSITTPTS	PSFPSPSFPSPSFPSPSFKASPSFKASPS	203					
Qy	60	P-PAPAPAKAPARELDRESGVENKLIVSG	FSPASTPLPDGTGENSTENNNAVTP	118					
Dd	204	PFKASPSFKASPARSPQSPSTPSPKAS	PVASPOOSPTSRPRSPTSPPTSP-	258					
Qy	119	VSOGSNSSLADNKAPP--PPVSSGEPT	145						
Dd	259	SPKASPPFSFGASPSLSRKVSSTPT	287						
RESULT 13									
Q39492	ID	Q39492	PRELIMINARY; PRT; 420 AA.						
IC	Q39492; QBS247;								
AD	01-MAY-2000 (tREMBUrel_13, Created)								
DT	01-OCT-2002 (tREMBUrel_22, Last sequence update)								
DD	01-MAR-2003 (tREMBUrel_23, Last annotation update)								
DE	CG15021 protein (RB1716sp).								
GN	CG15021.								
OC	Drosophila melanogaster (fruit fly).								
CC	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;								
CO	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;								



OC Ephydroidea; Drosophilidae; Drosophila.  
 RN NCBI\_TaxID=7227;  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anand S., Lewis S.E., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.A., Smit J.P., Yandell M.D., Zhang Q., Chen L.Y.,  
 RA Brandon R.C., Rogers Y., H.C. Blazer P.G., Champ M., Pfeiffer B.D.,  
 RA Wen K.H., Doyle C., Agayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
 RA April J.F., Agayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
 RA Ballew R.M., Basu P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Beeson K.Y., Borkov D., Bouch J., Brokstein P., Brotter P.,  
 RA Borkov D., Borkov D., Bouch J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busem D.A., Butler H., Cadiou E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahle C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mout S.M., Milshina N.V., Moberly C., Morris J., Moshrefi A.,  
 RA Mount S.W., Murphy B., Murphy L., Muzoy D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclebb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reineck K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Masarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhong M., Zhou X., Zhu S., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RT Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon K.Y., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champ M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett E., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frisse E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jallali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Paclebb J., Paragas V., Park S., Patel S., Pfeiffer B., Scheeler F.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of drosophila melanogaster genome";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Miera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frisse E., de Grey A., Harris N.,  
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of drosophila melanogaster genome";  
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]

RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RA Stapleton M., Brokstein P., Hong L., Agayani A., Carlson J.,  
 RA George R., Chavez C., Dorsett V., Dresnek D., Farfan D., Frisse E.,  
 RA Miranda A., Mungall C.J., Nunoo J., Paclebb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB003481; AAF47902.2; -;  
 DR EMBL; AY071124; AAL48746.1; -;  
 DR FlyBase; FBgn0035544; CG15021.  
 DR InterPro; IPR003882; Pistil\_extensin.  
 DR InterPro; IPR002965; P\_rich\_extensin.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 DR PRINTS; PR01218; PSTEXTENSIN.  
 SQ SEQUENCE 420 AA; 42947 MW; 96D62FFDC9F996E1 CRC64;  
 Query Match 15 43; Score 158.5; DB 5; Length 420;  
 Best Local Similarity 29.63; Pred No 0.00038;  
 Matches 42; Conservative 11; Mismatches 64; Indels 25; Gaps 3;  
 QY 33 KPFLPQPPAPANQDNSSQNTLQPTTPAPA--PKPAPPPPLDRSPGVENKLIPS 90  
 DB 71 PQTTPPPPPPP-----QTPPPAPRYSGPOTQPPPPPTPSAPAPPPPS 118  
 QY 91 VG-----SPASSTPLPDGTGNSTNNRNVTVSGSGNSSSADPKAPPPPPVS 139  
 DB 119 YGPQTTPPPPPPTPSAPAPPPSYGPPQTTPPPPPPTPSAPAPPPPPQ 178  
 QY 140 SGEPTTLGENDGLSQSLHR 161  
 DB 179 PPSPPSPQPGPEYLPDPQPKR 200  
 RESULT 14  
 ID Q41805 PRELIMINARY; PRT; 1188 AA.  
 AC Q41805;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Extensin-like protein precursor.  
 OS Zea mays (Maize protein precursor).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Zea.  
 ON NCBI\_TaxID=4577;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B73; TISSUE=Pollen;  
 RA Rubinstein A.L., Broadwater A.H., Lowrey K., Bedinger P.A.;  
 RT "Pex genes: pollen-specific genes with extensin-like domains";  
 RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z34465; CAA84230.1; -;  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR007090; LRR\_plant.  
 DR Pfam; PF00560; LRR\_3\_rich\_extensin.  
 DR PRINTS; PR01217; PRICHEXTENSIN.  
 DR PROSITE; PS50502; LRR\_PS; 2.  
 KW Signal.  
 FT SIGNAL. 1 27 POTENTIAL  
 SQ SEQUENCE 1188 AA; 120981 MW; 2C77C7F8D7130149 CRC64;

Query Match 15.4%; Score 158.5; DB 10; Length 1188;  
 Best Local Similarity 32.0%; Pred. No. 0.0011;  
 Matches 49; Conservative 17; Mismatches 68; Indels 19; Gaps 7;  
 QY 14 KTERSTAPLNTQISMLRNDKPLP-QPPAPANQNSQNTLQP---TPPIAPAPK 68  
 DB 961 KSSPPAPVNLPPPVKSSPPPTFVSSSPAPKSPPPPPVKSPPPPVSS 1020  
 QY 69 P-----AAPPLRLDRESGVENKLIIPVSPASSTPLP---PGTGNSTNNRAVTPV 119  
 DB 1021 PPPPVKSPPPAPVSSPPPVKS---PPPPAPVSSPPPPVKSPPPPAPVSSPPPVKSP 1077  
 QY 120 SQGSSNSSADP-KAPPPP-FVSSGEPPTLGENP 150  
 DB 1078 PPAPVSSPPPVKSPPPPVKSPPPPVKSPP 1110

## RESULT 15

P70433 PRELIMINARY; PRT; 802 AA.  
 AC P70433; P70430; P70431; P70432;  
 DT 01-FEB-1997 (TREMUR1. 02, Created)  
 DT 01-FEB-1997 (TREMUR1. 02, Last sequence update)  
 DT 01-MAR-2003 (TREMUR1. 23, Last annotation update)  
 DE ENABLED homolog.  
 GN ENAH OR MENA.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RW [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=97015079; PubMed=8861907;  
 RA Gertler F.B., Niebuhr K., Reinhard M., Wehland J., Soriano P.;  
 RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the  
 control of microfilament dynamics.";  
 RL Cell 87:227-239(1996).  
 CC -!- FUNCTION: MAY BE INVOLVED IN MICROFILAMENT ASSEMBLY AND CELL  
 MOTILITY. INDUCES THE FORMATION OF F-ACTIN RICH OUTGROWTHS IN  
 FIBROBLASTS. BINDS PROFILIN.  
 CC -!- SUBCELLULAR LOCATION: LOCALIZED TO FOCAL ADHESIONS AND, TO A  
 LESSER EXTENT, LEADING EDGES AND STRESS FIBERS.  
 CC -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS, MENA, MENA+, MENA++ AND MENA+++  
 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 DR ENBL; U72520; AAC52863.1; -;  
 DR ENBL; U72521; AAC52864.1; -;  
 DR ENBL; U72522; AAC52865.1; -;  
 DR ENBL; U72523; AAC52866.1; -;  
 DR MGD; MGI:106360; Enah.  
 DR InterPro; IPR000697; EVH1.  
 DR InterPro; IPR000156; Ran\_BP1.  
 DR InterPro; IPR001960; WH1.  
 DR Pfam; PF00568; WH1; 1.  
 DR SMART; SMO0461; RanBd; 1.  
 DR SMART; SMO0461; WH1; 1.  
 KW Alternative splicing; Phosphorylation.  
 FT DOMAIN 442 464  
 FT POLY-PRO.  
 FT DOMAIN 542 552  
 FT POLY-PRO.  
 FT DOMAIN 562 574  
 FT POLY-PRO.  
 FT DOMAIN 578 589  
 FT POLY-PRO.  
 FT DOMAIN 593 605  
 FT POLY-PRO.  
 FT DOMAIN 686 689  
 FT POLY-ARG.  
 FT MOD\_RES 255 255  
 FT PHOSPHORYLATION (BY CAPK AND CGPK) (BY  
 SIMILARITY)  
 FT MOD\_RES 637 637  
 FT PHOSPHORYLATION (BY CAPK AND CGPK) (BY  
 SIMILARITY).  
 FT VARSPLIC 117 131  
 FT MISSING (IN ISOFORM MENA+).  
 FT VARSPLIC 117 135  
 FT MISSING (IN ISOFORM MENA AND ISOFORM  
 MENA+).  
 FT VARSPLIC 132 135  
 FT CIFS -> VFYL (IN ISOFORM MENA+++).  
 FT VARSPLIC 259 500  
 FT MISSING (IN ISOFORM MENA).  
 SQ SEQUENCE 802 AA; 85844 MW; 592BB975EE20F77F CRC64;

Query Match 15.4%; Score 158; DB 11; Length 802;  
 Best Local Similarity 26.3%; Pred. No. 0.00079;  
 Matches 59; Conservative 15; Mismatches 62; Indels 88; Gaps 10;  
 QY 13 NKTERSTAPLNT-----QISALRNDP----- 33  
 DB 348 NKNSRPSSPVNTSSQPPAAKSCAWPTSNFSLPPSPPTIMISSPPGKATGPRVLPVCVS 407  
 QY 34 KPLPQPPAPAPANQDN-----SSQNTRLQPTPTPIAPAPKPAAPRPL----- 76  
 DB 408 SPVQMPSPPTAPNGSLDSVTVFVSPPPTSCPAAPPPPPPPPPPPPPPPPPPPPPPL 467  
 QY 77 -----DRESGVENKLI-----PSVGSPASS-TLPPD-GTGNSTNNRAV. 116  
 DB 468 LSHCGSQASPPPGTFLASTPSSKPSVLPFSAGAPASATETPLNFELODSSASEPQLQAS 527  
 QY 117 -----TPVSGGSSNSADPKAPPPPPVSSGE-----PPTLGENP 150  
 DB 528 QPAESPTPG--LVLGPPAPPPPPPLPSGPAYASALPPPPGPP 569

Search completed: November 13, 2003, 09:27:33  
 Job time : 66.4086 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 / Search time 14.2996 Seconds  
(without alignments)  
388.502 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQBLEHRSRLQTRDQRLPFDEKFTGA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	1426	23	Human legless homo
2	178	100.0	1435	22	Human BCL9 homolog
3	109	61.2	320	23	Mouse beta-catenin
4	109	61.2	1494	23	Mouse beta-catenin
5	107	60.1	738	23	Mouse beta-catenin
6	107	60.1	1115	23	Human legless homo
7	66	37.1	1429	22	Aspergillus fumig
8	66	37.1	1464	23	D. melanogaster lg
9	57.5	32.3	603	24	Aspergillus fumiga

10	57.5	32.3	618	24	ABJ26453	Aspergillus fumiga
11	56	31.5	1294	22	ABE63502	Drosophila melanog
12	55	30.9	294	22	AAB95073	Human protein sequ
13	55	30.9	390	23	ABG96285	Human ovarian can
14	55	30.9	433	22	AA663851	Human ovarian can
15	55	30.9	433	22	AA663852	Amino acid sequenc
16	55	30.9	433	22	AA663852	Amino acid sequenc
17	55	30.9	464	22	AA663852	Human GTP-binding
18	55	30.9	464	22	AA663852	Human protein sequ
19	55	30.9	464	22	AA663852	Human ovarian can
20	55	30.9	464	22	AA663852	Human ovarian can
21	55	30.9	718	23	ABG96284	Human ovarian can
22	54.5	30.6	174	22	ABE6346	Murine BHLH trans
23	54.5	30.6	174	22	ABE6346	Murine BHLH trans
24	53	29.8	819	22	AAU69744	Thermus thermophil
25	52.5	29.5	181	22	ABE6345	Human BHLH trans
26	52.5	29.5	181	22	ABE6345	Human BHLH trans
27	52	29.2	959	21	AA53051	Human secreted pro
28	51.5	28.9	210	23	ABP43969	Adenomatous poly
29	51	28.7	329	21	AAJ31171	Arabidopsis thalia
30	51	28.7	434	22	AAU33491	Enterococcus faeca
31	51	28.7	440	23	ABP4770	Protein #16 relate
32	51	28.7	448	22	AAU35058	Enterococcus faeca
33	51	28.7	500	22	AA89346	Human secreted pro
34	51	28.7	1464	22	ABE71111	Drosophila melanog
35	51	28.7	1755	20	AA411139	Mouse mammary tumo
36	50.5	28.4	96	22	ABE69662	Drosophila melanog
37	50.5	28.4	479	23	ABP43965	Unidentified prote
38	50.5	28.4	675	21	AAV54052	A variant of an an
39	50.5	28.4	675	21	AAV54053	Angiogenesis-as
40	50.5	28.4	1183	22	ABE58769	Drosophila melanog
41	50	28.1	187	21	AA838555	Human secreted pro
42	50	28.1	221	21	AA807850	Amino acid sequenc
43	50	28.1	223	22	ABE62235	Glycine max gruat
44	50	28.1	275	21	AA824369	Bacillus subtilis
45	50	28.1	342	20	AA442781	Human neuronal imm

#### ALIGNMENTS

RESULT 1  
AAB71229  
ID AAB71229 standard; Protein; 1426 AA.

AC AAB71229;

DT 18-NOV-2002 (first entry)

XX Human legless homologue lgs/bcl9 protein.

XX Legless; human; lgs; Wnt/Wingless signaling pathway; Mot; Mg;  
XX tissue proliferation; tumour; cytosatic; cellular disorder; colon;  
XX blood disorder; cancer; breast; head and neck cancer; thyroid;  
XX medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX Homo sapiens.

XX US2002086986-A1.

XX 04-JUL-2002.

XX 27-JUL-2001; 2001US-0915543.

XX 28-JUL-2000; 2000US-221502P.

XX (BASL/) BASLER K.

XX (BRUN/) BRUNNER E.

XX (FROE/) FROESCH B.

XX (KRAM/) KRAMPS T.

XX (PETE/) PETER O.

XX Basler K, Brunner E, Froesch B, Kramps T, Peter O;

```

XX WPI; 2002-635689/68.
XX N-PSDB; AAF88467.
XX
PT Novel polypeptide useful in therapeutic method for treating disorders
PT of cell fate such as cell differentiation or cell proliferation
XX
PS Example II; Fig 8B; 41pp; English.
XX
XX This invention describes a novel polypeptide sharing one or more
XX homologous amino acid domains with the legless (lgs) protein, a
XX downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway
XX involved in the formation and maintenance of spatial arrangements
XX and proliferation of tissues during development, and in the formation
XX and growth of many human tumours. The products of the invention have
XX cytoskeletal activity and can be used to treat cellular disorders, blood
XX disorders and cancers caused by over-stimulation of the Wnt pathway,
XX where the cancerous condition is colon, breast, head and neck, brain,
XX thyroid, medulloblastoma or skin cancer. The product could also be used
XX to promote tissue regeneration and repair. This sequence represents the
XX human legless (lgs) protein homologue lgs/bcl9 described in the
XX disclosure of the invention.
XX
SQ Sequence 1426 AA;
Query Match 100.0%; Score 178; DB 23; Length 1426;
Best Local Similarity 100.0%; Pred. No. 8.6e-16;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DLSQQLHRRRSLSQTLRDIQRMFLPDPKEFTGA 35
DB 349 DLSQQLHRRRSLSQTLRDIQRMFLPDPKEFTGA 383
RESULT 2
ABE11808
ID ABE11808 standard; peptide; 1435 AA.
XX
XX ABE11808;
XX
DT 11-JAN-2002 (first entry)
DE Human BCL9 homologue, SEQ ID NO:2178.
XX
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX myeloid cell disorder; lymphoid cell disorder; arthritis;
XX chronic inflammatory condition; proliferative retinopathy;
XX atherosclerosis; coronary heart disease; arterial ischaemia;
XX bone disorder; osteoporosis; vascular growth disorder;
XX tissue regeneration; wound healing; infection; immune disorder;
XX cell culture; drug screening; gene therapy; anti-inflamatory;
XX antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
XX cytoskeletal; osteopathic; vasotropic; cardiac; virucide; antibacterial;
XX antifungal; vulnery; antiulcer.
XX
OS Homo sapiens.
XX
DN WO200157188-A2.
XX
PD 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-0503800.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C. Drmanac RT;
XX

```

---

```

DR WPI; 2001-457740/49.
DR N-PSDB; ABA09052.
XX
PT Human proteins and DNA encoding sequences useful for preventing,
PT treating or ameliorating a medical condition in a mammalian subject
PT e.g. arthritis and cancer -
XX
PS Claim 20; Page 256-257; 1963pp; English.
XX
XX Sequences ABE10981-ABE12330 represent 1350 novel human polypeptides, and
XX sequences ABA09225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX antibodies against the polypeptides, methods of detecting the nucleotides
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, and hence
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoiesis regulatory activity; tissue growth activity;
XX immunomodulatory activity; activin- or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; or may be
XX involved in oncogenesis, cancer cell proliferation or metastasis.
XX Depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX vascular growth. Polypeptides involved with tissue regeneration and
XX repair (or nucleic acids encoding them) may be used to promote wound
XX healing (e.g., of burns, incisions and ulcers), while those with
XX immunomodulatory activities may be used in the treatment of viral,
XX bacterial and fungal infections in addition to immune disorders.
XX Polypeptides with growth factor activity may be used in cell cultures to
XX promote cell growth. For example, such polypeptides may be used to
XX manipulate stem cells in culture to give rise to neuroepithelial cells
XX that can be used to augment or replace cells damaged by illness.
XX autoimmune disease or accidental damage. The polypeptides and nucleotides
XX may also be used in the diagnosis of the above conditions, and in drug
XX screening techniques. The present sequence represents a novel human
XX polypeptide of the invention.
XX
SQ Sequence 1435 AA;
Query Match 100.0%; Score 178; DB 22; Length 1435;
Best Local Similarity 100.0%; Pred. No. 8.7e-16;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DLSQQLHRRRSLSQTLRDIQRMFLPDPKEFTGA 35
DB 389 DLSQQLHRRRSLSQTLRDIQRMFLPDPKEFTGA 423
RESULT 3
AAU78461
ID AAU78461 standard; Protein; 320 AA.
XX
XX AAU78461;
XX
DT 02-JUL-2002 (first entry)
XX
XX Mouse beta-catenin nuclear localised protein #2.
XX
DE Mouse; beta-catenin nuclear localised protein; cancer;
XX gene therapy; EST; expressed sequence tag.
XX
XX Mus musculus.
XX

```

PN WO200224738-A1.  
 XX 28-MAR-2002.  
 XX 19-SEP-2001; 2001WO-JP08140.  
 XX 22-SEP-2000; 2000JP-0287876.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX Akiyama T, Adachi S;  
 XX WPI; 2002-330014/36.  
 XX N-PSDB; ABK47632.  
 XX New beta-catenin nuclear localised protein for diagnosis and treatment  
 PT of diseases associated with nuclear localisation of beta-catenin e.g.  
 PT cancer -  
 XX  
 XX Claim 2; Page 91-92; 113pp; Japanese.  
 XX The invention relates to a beta-catenin nuclear localised protein  
 CC and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC mouse beta-catenin nuclear localised protein #2.  
 XX Sequence 320 AA;  
 XX  
 XX Query Match 61.2%; Score 109; DB 23; Length 320;  
 XX Best Local Similarity 84.0%; Pred. No. 6.3e-07;  
 XX Matches 21; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 DGLSQQLHRRSLQTLRDIOQML 25  
 DB 150 EGLSKQLHRRSLQTLRDIERLL 174  
 RESULT 4  
 AAU78460  
 ID AAU78460 standard; Protein; 1494 AA.  
 AC AAU78460;  
 XX 02-JUL-2002 (first entry)  
 XX Mouse beta-catenin nuclear localised protein.  
 XX Mouse; beta-catenin nuclear localised protein; cancer;  
 KW gene therapy; EST; expressed sequence tag.  
 KW Mus musculus.  
 OS WO200224738-A1.  
 XX 28-MAR-2002.  
 XX 19-SEP-2001; 2001WO-JP08140.  
 XX 22-SEP-2000; 2000JP-0287876.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX Akiyama T, Adachi S;  
 XX WPI; 2002-330014/36.  
 XX N-PSDB; ABK47631.  
 XX New beta-catenin nuclear localised protein for diagnosis and treatment  
 PT of diseases associated with nuclear localisation of beta-catenin e.g.  
 PT cancer -  
 XX

PS Claim 1; Page 81-88; 113pp; Japanese.  
 XX The invention relates to a beta-catenin nuclear localised protein  
 CC and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC mouse beta-catenin nuclear localised protein.  
 XX Sequence 1494 AA;  
 XX  
 XX Query Match 61.2%; Score 109; DB 23; Length 1494;  
 XX Best Local Similarity 84.0%; Pred. No. 3.5e-06;  
 XX Matches 21; Conservative 4; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 DGLSQQLHRRSLQTLRDIOQML 25  
 DB 394 EGLSKQLHRRSLQTLRDIERLL 418  
 RESULT 5  
 AAU78463  
 ID AAU78463 standard; Protein; 738 AA.  
 AC AAU78463;  
 XX 02-JUL-2002 (first entry)  
 XX Human beta-catenin nuclear localised protein #2.  
 XX Mouse; beta-catenin nuclear localised protein; cancer;  
 KW gene therapy; EST; expressed sequence tag.  
 KW Homo sapiens.  
 OS WO200224738-A1.  
 XX 28-MAR-2002.  
 XX 19-SEP-2001; 2001WO-JP08140.  
 XX 22-SEP-2000; 2000JP-0287876.  
 XX (KYOW ) KYOWA HAKKO KOGYO KK.  
 XX Akiyama T, Adachi S;  
 XX WPI; 2002-330014/36.  
 XX N-PSDB; ABK47638.  
 XX New beta-catenin nuclear localised protein for diagnosis and treatment  
 PT of diseases associated with nuclear localisation of beta-catenin e.g.  
 PT cancer -  
 XX  
 XX Claim 8; Page 102-105; 113pp; Japanese.  
 XX The invention relates to a beta-catenin nuclear localised protein  
 CC and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC human beta-catenin nuclear localised protein #2.  
 XX Sequence 738 AA;  
 XX  
 XX Query Match 60.1%; Score 107; DB 23; Length 738;  
 XX Best Local Similarity 87.5%; Pred. No. 3e-06;  
 XX Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 GLSQQLHRRSLQTLRDIOQML 25  
 DB 1 GLSKQLHRRSLQTLRDIERLL 24

## RESULT 6

AAAB71230  
ID AAB71230 standard; Protein; 1115 AA.

XX AC AAB71230;  
XX DT 18-NOV-2002 (first entry)

XX XX Human legless homologue higs-1 partial protein.

XX DE Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
XX KW tissue proliferation; tumour; cytoskeletal; cellular disorder; colon;  
XX KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
XX KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX OS Homo sapiens.  
XX XX US2002086986-A1.  
XX PN 04-JUL-2002.  
XX PD 27-JUL-2001; 2001US-0915543.  
XX PF 28-JUL-2000; 2000US-221502P.

XX PR (BASL/) BASLER K.  
XX PA (BRUN/) BRUNNER E.  
XX PA (FROE/) FROESCH B.  
XX PA (KRAM/) KRAMPS T.  
XX PA (PETE/) PETER O.

XX PI Basler K, Brunner E, Froesch B, Kramps T, Peter O;  
XX XX WPI: 2002-635685/68.  
XX DR N-PSDB; AAF88468.

XX XX Novel polypeptide useful in therapeutic method for treating disorders  
XX PT of cell fate such as cell differentiation or cell proliferation -  
XX XX Example II; Fig 10B; 41pp; English.

XX CC This invention describes a novel polypeptide sharing one or more  
XX CC homologous amino acid domains with the legless (lgs) protein, a  
XX CC downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
XX CC involved in the formation and maintenance of spatial arrangements  
XX CC and proliferation of tissues during development, and in the formation  
XX CC and growth of many human tumours. The products of the invention have  
XX CC cytoskeletal activity and can be used to treat cellular disorders, blood  
XX CC disorders and cancers caused by over-stimulation of the Wnt pathway,  
XX CC where the cancerous condition is colon, breast, head and neck, brain,  
XX CC thyroid, medulloblastoma or skin cancer. The product could also be used  
XX CC to promote tissue regeneration and repair. This sequence represents the  
XX CC human legless (lgs) protein homologue higs-1 described in the  
XX CC disclosure of the invention.

XX SQ Sequence 1115 AA;  
XX Query Match 60.1%; Score 107; DB 23; Length 1115;  
XX Best Local Similarity 87.5%; Pred. No. 4.8e-06;  
XX Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLSQQLHREHRSLSQTLDRQLRWL 25

DB 76 GLSQQLHREHRSLSQTLDRQLRWL 99

## RESULT 7

ABBS8779  
ID ABBS8779 standard; Protein; 1429 AA.

XX AC ABBS8779;  
XX XX

DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 3129.  
XX AC AAB71230;  
XX KW Drosophila; developmental biology; cell signalling; insecticide;  
XX KW pharmaceutical.

XX OS Drosophila melanogaster.

XX XX WO200171042-A2.

XX DE 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX XX (PEXE ) PE CORP NY.

XX XX Venter JC, Adams M, Li PWD, Myers EW;  
XX XX WPI: 2001-656860/75.  
XX DR N-PSDB; ABL02882.

XX XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX PT genes from Drosophila and for elucidating cell signalling and cell-cell  
XX PT interactions -  
XX XX Disclosure; SEQ ID NO 3129; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent  
XX CC capable of detecting 1000 or more genes from Drosophila. The invention is  
XX CC useful in developmental biology and in elucidating cell signalling and  
XX CC cell-cell interactions in higher eukaryotes for the development of  
XX CC insecticides, therapeutics and pharmaceutical drugs. The invention  
XX CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA  
XX CC sequences (ABU16176-ABU30511) and the encoded proteins  
XX CC (ABBS8779-ABBS8779).

XX CC The sequence data for this patent did not form part of the printed  
XX CC specification, but was obtained in electronic format directly from WIPO  
XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 1429 AA;

XX Query Match 37.1%; Score 66; DB 22; Length 1429;  
XX Best Local Similarity 31.4%; Pred. No. 3.1;  
XX Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 DGLSQQLHREHRSLSQTLDRQLRWLPPDKEFTGA 35

DB 520 ENLTPQQRHREBQLAKIKKNNQFLFENENSVGA 554

## RESULT 8

AAAB71228  
ID AAB71228 standard; Protein; 1464 AA.

XX AC AAB71228;  
XX XX 18-NOV-2002 (first entry)

XX DE D. melanogaster lgs protein.

XX KW Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
XX KW tissue proliferation; tumour; cytoskeletal; cellular disorder; colon;  
XX KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
XX KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.

XX OS Drosophila melanogaster.

XX XX US2002086986-A1.

PD 04-JUL-2002.  
PF 27-JUL-2001; 2001US-0915543.  
XX 28-JUL-2000; 2000US-221502P.  
XX (BASL/) BASLER K.  
PA (BRUN/) BRUNNER E.  
PA (FROE/) FROESCH B.  
PA (KRAM/) KRAMP T.  
PA (PETE/) PETER O.  
XX  
XX Basler K, Brunner E, Froesch B, Kramps T, Peter O;  
XX WPI; 2002-635689/69.  
XX N-PSDB; AAF88466.  
XX  
XX Novel polypeptide useful in therapeutic method for treating disorders  
XX of cell fate such as cell differentiation or cell proliferation  
XX  
XX Example II; Fig 2; 41pp; English.  
XX  
XX This invention describes a novel polypeptide sharing one or more  
XX homologous amino acid domains with the legless (lgs) protein, a  
XX downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
XX involved in the formation and maintenance of spatial arrangements  
XX and proliferation of tissues during development, and in the formation  
XX and growth of many human tumors. The products of the invention have  
XX cytostatic activity and can be used to treat cellular disorders, blood  
XX disorders and cancers caused by over-stimulation of the Wnt pathway,  
XX the cancerous condition is colon, breast, head and neck, brain,  
XX thyroid medullary carcinoma or skin cancer. The product could also be used  
XX to promote tissue regeneration and repair. This sequence represents the  
XX Drosophila melanogaster (fruitfly) legless (lgs) protein described in  
XX the disclosure of the invention.  
XX  
XX Sequence 1464 AA;  
SQ  
Query Match 37.1%; Score 66; DB 23; Length 1464;  
Best Local Similarity 31.4%; Pred. No. 3.2;  
Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
QY 1 DGLSQEQLHREHRSLSQTLRDQIQLMFLPDEKEFTGA 35  
DB 515 ENLTPQQRHREELQAKIKKQNFLEPENSUGA 549  
RESULT 9  
ABJ25853  
ID ABJ25853 standard; Protein; 603 AA.  
AC ABJ25853;  
XX  
XX 16-APR-2003 (first entry)  
XX  
XX Aspergillus fumigatus essential gene protein #511.  
XX  
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
XX cancer; contamination; biofilm; antibody; immune response.  
XX  
XX Aspergillus fumigatus.  
XX  
XX WO200286090-A2.  
XX  
XX 31-OCT-2002.  
XX  
XX 23-APR-2002; 2002WO-US131142.  
XX  
XX 23-APR-2001; 2001US-285697P.  
XX 27-APR-2001; 2001US-287066P.  
XX 05-JUN-2001; 2001US-295890P.  
XX 09-JUL-2001; 2001US-303899P.  
XX 31-AUG-2001; 2001US-316362P.  
XX

XX  
XX (ELIT-) ELITRA PHARM INC.  
XX Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;  
XX WPI; 2003-093124/08.  
XX  
XX New purified or isolated nucleic acids of essential genes of  
XX Aspergillus fumigatus, useful for treating or preventing infections by  
XX A. fumigatus, or for treating a non-infectious disease in a subject  
XX e.g. cancer  
XX  
XX Disclosure; Page -: 175pp; English.  
XX  
XX The invention relates to novel purified or isolated nucleic acids of  
XX essential genes of Aspergillus fumigatus. The isolated nucleic acids of  
XX the invention are used to treat or prevent infections by a pathogenic  
XX organism such as A. fumigatus, to treat a non-infectious disease in a  
XX subject (e.g. cancer), to prevent or contain contamination of an object  
XX by A. fumigatus, or to prevent or inhibit formation on a surface of a  
XX biofilm comprising A. fumigatus. The polynucleotides are useful for  
XX expressing recombinant protein for characterization, screening or  
XX therapeutic use, as markers for host tissues in which the pathogenic  
XX organisms invade or reside, for comparing with the DNA sequence of A.  
XX fumigatus to identify duplicated genes or paralogues having the same or  
XX similar biochemical activity and/or function, for comparing with DNA  
XX sequences of other related or distant pathogenic organisms to identify  
XX potential orthologous essential or virulence genes, for selecting and  
XX making oligomers for attachment to a nucleic acid array for examination  
XX of expression patterns, for raising anti-protein antibodies, as an  
XX antigen to raise anti-DNA antibodies or to elicit another immune  
XX response and for identifying polynucleotides encoding the other protein  
XX with which binding occurs or to identify inhibitors of the binding  
XX interaction. The polypeptides may be used to raise antibodies or to  
XX elicit immune response, as a reagent in assays designed to quantitatively  
XX determine levels of the protein in biological fluids, as a marker for  
XX host tissues in which pathogenic organism invade or reside, and to  
XX isolate correlative receptors or ligands in the case of virulence  
XX factors. This sequence represents a protein of one of the essential genes  
XX of Aspergillus fumigatus of the invention.  
XX  
XX Sequence 603 AA;  
SQ  
Query Match 32.3%; Score 57.5; DB 24; Length 603;  
Best Local Similarity 31.4%; Pred. No. 18;  
Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;  
QY 1 DGLSQEQLHREHRSLSQTLRDQIQLMFLPDEKEFTGA 35  
DB 45 DGVETEKIREND--EVEKKLERMLFGDDGFVGA 76  
RESULT 10  
ABJ26453  
ID ABJ26453 standard; Protein; 618 AA.  
AC ABJ26453;  
XX  
XX 16-APR-2003 (first entry)  
XX  
XX Aspergillus fumigatus essential gene protein #1111.  
XX  
XX Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;  
XX cancer; contamination; biofilm; antibody; immune response.  
XX  
XX Aspergillus fumigatus.  
XX  
XX WO200286090-A2.  
XX  
XX 31-OCT-2002.  
XX  
XX 23-APR-2002; 2002WO-US131142.  
XX

PR	23-APR-2001; 2001US-285697P.
PR	27-APR-2001; 2001US-287066P.
PR	05-JUN-2001; 2001US-295890P.
PR	09-JUL-2001; 2001US-304899P.
PR	31-AUG-2001; 2001US-316362P.
XX	(ELIT-) ELITRA PHARM INC.
PA	Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
PI	WPI; 2003-093124/08.
DR	New purified or isolated nucleic acids of essential genes of Aspergillus fumigatus, useful for treating or preventing infections by A. fumigatus, or for treating a non-infectious disease in a subject e.g. cancer
PT	Disclosure; Page -: 175pp; English.
XX	The invention relates to novel purified or isolated nucleic acids of essential genes of Aspergillus fumigatus. The isolated nucleic acids of the invention are used to treat or prevent infections by a pathogenic organism such as A. fumigatus, to treat a non-infectious disease in a subject (e.g. cancer), to prevent or contain contamination of an object by A. fumigatus, or to prevent or inhibit formation on a surface of a biofilm comprising A. fumigatus. The polynucleotides are useful for expressing recombinant protein for characterization, screening or therapeutic use, as markers for host tissues in which the pathogenic organisms invade or reside, for comparing with the DNA sequence of A. fumigatus to identify duplicated genes or paralogues having the same or similar biochemical activity and/or function, for comparing with DNA sequences of other related or distant pathogenic organisms to identify potential orthologous essential or virulence genes, for selecting and making oligomers for attachment to a nucleic acid array for examination of expression patterns, for raising anti-protein antibodies, as an antigen to raise anti-DNA antibodies or to elicit another immune response, and for identifying polynucleotides encoding the other protein with which binding occurs or to identify inhibitors of the binding interaction. The polypeptides may be used to raise antibodies or to elicit immune response, as a reagent in assays designed to quantitatively determine levels of the protein in biological fluids, as a marker for host tissues in which pathogenic organisms invade or reside, and to isolate correlative receptors or ligands in the case of virulence factors. This sequence represents a protein of one of the essential genes of Aspergillus fumigatus of the invention.
CC	Sequence 618 AA;
SQ	Query Match 32.3%; Score 57.5; DB 24; Length 618; Best Local Similarity 31.4%; Pred. No.19; Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1
QY	1 DLGSQLHRSRLQTRDIQRMLFPDEKFTGA 35    ::: : : :      ::  45 DGVTETIREKD---EVEKKLRMLFGDGEFVGA 76
DB	RESULT 11 ID AB65302 ID AB65302 standard; Protein; 1294 AA. AC AB65302; AC AB65302; DT 26-MAR-2002 (first entry) DX Drosophila melanogaster polypeptide SEQ ID NO 17298. DE Drosophila; developmental biology; cell signalling; insecticide; XW pharmaceutical. KW Drosophila melanogaster. OS Drosophila melanogaster. SN WO200171042-A2.



PI Ishii S, Sugiyama T, Wakamatsu A, Nogai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 FT full-length cDNAs -  
 XX  
 XX Claim 8; SEQ ID 16943; 2537pp + CD ROM; English.  
 XX  
 XX The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dr primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH97446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 XX Sequence 294 AA;  
 SQ  
 Query Match 30.9%; Score 55; DB 22; Length 294;  
 Best Local Similarity 52.2%; Pred. No. 18;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQEQLHERSLSQTLDRQRM 24  
 DB 224 GLRTGGLFRSASVQTVREIQL 246  
 RESULT 13  
 ID ABG96285  
 AC ABG96285;  
 XX  
 XX 11-DEC-2002 (first entry)  
 DT  
 XX Human ovarian cancer marker M355.  
 DE  
 XX Human; ovarian cancer; marker; cancer; familial history; brain disorder;  
 KW central nervous system disorder; bacterial meningitis; viral meningitis;  
 KW Alzheimer's disease; Parkinson's disease; cerebral oedema; hydrocephalus;  
 KW brain herniation; inflammation; encephalitis; testicular disorder;  
 KW non-tuberculous granulomatous orchitis; connective tissue disorder;  
 KW heart disorder; ischaemic heart disease; atherosclerosis; neoplasm;  
 KW histological type; carcinogenic; ovarian cancer marker.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200271928-A2.  
 PN  
 XX 19-SEP-2002.  
 PD  
 XX 14-MAR-2002; 2002WO-US07826.  
 PF  
 XX 14-MAR-2001; 2001US-276025P.  
 PR  
 XX 14-MAR-2001; 2001US-276026P.

PR 10-AUG-2001; 2001US-311732P.  
 PR 19-SEP-2001; 2001US-323580P.  
 PR 26-SEP-2001; 2001US-324967P.  
 PR 26-SEP-2001; 2001US-325102P.  
 PR 26-SEP-2001; 2001US-325149P.  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA  
 XX Monahan JE, Ganavarapu M, Hoersch S, Kamatkar S, Kovatis SG;  
 PI Meyers RE, Morrissey MP, Olandt FO, Sen A, Visey PO, Mills GB;  
 PI East RC, Lu K, Schmandt RE, Zhao X, Glatt K;  
 XX WPI; 2002-723277/78.  
 XX N-PSDB; ABS76377.  
 DR  
 DR Assessing whether a patient is afflicted with ovarian cancer, useful in  
 PT assessing the stage or progression of the disease, comprises comparing  
 PT the expression level of a cancer marker in a sample from a patient and  
 PT from a non cancer patient -  
 FT  
 XX Disclosure; Page 134-135; 481pp; English.  
 PS  
 XX The present invention relates to a new method for assessing whether a  
 CC patient is afflicted with ovarian cancer. The method involves comparing  
 CC the expression level of a marker in a patient sample and the normal level  
 CC of expression of the marker in a control non-ovarian cancer sample, where  
 CC the marker is selected from 363 cancer markers described in the  
 CC specification. The method of the invention is useful in diagnosing or  
 CC characterizing cancer, in detecting the presence of cancer as early as  
 CC possible, and the recurrence of ovarian cancer. The method may also be of  
 CC particular use with patients having an enhanced risk of developing  
 CC ovarian cancer (e.g. patients having a familial history of ovarian  
 CC cancer). The cancer markers may be used in the management and treatment  
 CC of e.g. brain and central nervous system disorders (e.g. bacterial and  
 CC viral meningitis, Alzheimer's disease or Parkinson's disease), brain  
 CC disorders (e.g. cerebral oedema, hydrocephalus or brain herniations),  
 CC inflammations (e.g. bacterial or viral meningitis or encephalitis),  
 CC testicular disorders (e.g. non-tuberculous granulomatous orchitis),  
 CC connective tissue disorders, or heart disorders (e.g. ischaemic heart  
 CC disease or atherosclerosis). The compositions and methods may also be  
 CC used in assessing the histological type of neoplasm associated with  
 CC ovarian cancer, monitoring the progression of ovarian cancer,  
 CC determining whether ovarian cancer has metastasized or is likely to  
 CC metastasize, selecting a composition for inhibiting ovarian cancer,  
 CC assessing the ovarian carcinogenic potential of a compound, or  
 CC inhibiting ovarian cancer or at risk of developing ovarian cancer. The  
 CC present amino acid sequence represents one of the ovarian cancer markers  
 CC described in the invention.  
 XX  
 XX Sequence 390 AA;  
 SQ  
 Query Match 30.9%; Score 55; DB 23; Length 390;  
 Best Local Similarity 52.2%; Pred. No. 25;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQEQLHERSLSQTLDRQRM 24  
 DB 181 GLRTGGLFRSASVQTVREIQL 203  
 RESULT 14  
 ID ABG63851  
 AC ABG63851 standard; Protein; 433 AA.  
 XX  
 XX ABG63851;  
 XX  
 XX 29-OCT-2001 (first entry)  
 DT  
 XX Amino acid sequence of human GTPase activating protein GTPAP1.  
 DE  
 XX GTPase activating protein; GTPAP1; cell signalling; immune disorder;  
 KW cell proliferative disorder; cancer; colon cancer; arteriosclerosis;  
 KW diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.

XX Homo sapiens.  
 XX Key Location/Qualifiers  
 XX Modified-site 60 /note= "potential protein kinase phosphorylation site for tyrosine kinase"  
 FT Modified-site 101 /note= "potential protein kinase phosphorylation site for tyrosine kinase"  
 FT Modified-site 129 /note= "potential protein kinase phosphorylation site for protein kinase C"  
 FT Modified-site 169 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Domain 210..235 /note= "GTPase-activator domain"  
 FT Domain 310..350 /note= "GTPase-activator domain"  
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for protein kinase C"  
 FT Modified-site 292 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Modified-site 297 /note= "potential protein kinase phosphorylation site for protein kinase C"  
 FT Modified-site 309 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Domain 310..350 /note= "GTPase-activator domain"  
 FT Modified-site 315 /note= "potential protein kinase phosphorylation site for tyrosine kinase"  
 FT Modified-site 338 /note= "potential N-linked glycosylation site"  
 FT Modified-site 382 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 XX WO200161010-A2.  
 XX 23-AUG-2001.  
 XX 15-FEB-2001; 2001WO-US05075.  
 XX 18-FEB-2000; 2000US-0507765.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Klingler TM, Stewart EA, Yue H, Baughn MR;  
 DR WPI; 2001-522598/57.  
 DR N-PSDB; AAH74978.  
 XX A substantially purified GTPase activating protein useful for treating or preventing cell signalling, immune and cell proliferative disorders, including cancer especially colon cancer -  
 XX Claim 14; Fig 1A-E; 78pp; English.  
 XX The present sequence represents a human GTPase activating protein, designated GTPAP1. GTPAP polynucleotides and polypeptides are useful for treating or preventing a disease or condition associated with altered expression of GTPase activating proteins, especially cell signalling, immune and cell proliferative disorders, including cancer especially colon cancer. Examples of diseases treatable include

CC arteriosclerosis, diabetes, psoriasis, hepatitis and multiple sclerosis.  
 CC GTPAP1 polynucleotides are also useful for gene therapy treatments of the diseases.  
 XX Sequence 433 AA;  
 XX Query Match 30.9%; Score 55; DB 22; Length 433;  
 XX Best Local Similarity 52.2%; Pred No. 28;  
 XX Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQPLEHRSLSLOTLDIORM 24  
 DB 224 GLRTEGLFRSASVQTVREIQL 246  
 RESULT 15  
 AAG63852  
 ID AAG63852 standard; Protein; 433 AA.  
 XX  
 AC AAG63852;  
 XX  
 DT 29-OCT-2001 (first entry)  
 XX  
 DE Amino acid sequence of human GTPase activating protein GTPAP2.  
 XX  
 KW GTPase activating protein; GTPAP2; cell signalling; immune disorder; cell proliferative disorder; cancer; colon cancer; arteriosclerosis; diabetes; psoriasis; hepatitis; multiple sclerosis; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT Modified-site 60 /note= "potential protein kinase phosphorylation site for tyrosine kinase"  
 FT Modified-site 101 /note= "potential protein kinase phosphorylation site for tyrosine kinase"  
 FT Modified-site 129 /note= "potential protein kinase phosphorylation site for protein kinase C"  
 FT Modified-site 169 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Domain 210..235 /note= "GTPase-activator domain"  
 FT Domain 310..350 /note= "GTPase-activator domain"  
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Modified-site 239 /note= "potential protein kinase phosphorylation site for protein kinase C"  
 FT Modified-site 292 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Modified-site 297 /note= "potential protein kinase phosphorylation site for protein kinase C"  
 FT Modified-site 309 /note= "potential protein kinase phosphorylation site for casein kinase II"  
 FT Domain 310..350 /note= "GTPase-activator domain"  
 FT Modified-site 315 /note= "potential protein kinase phosphorylation site for tyrosine kinase"  
 FT Modified-site 338 /note= "potential N-linked glycosylation site"  
 FT Modified-site 382 /note= "potential protein kinase phosphorylation site for casein kinase II"

```

XX WO200161010-A2.
PN
XX
XX 23-AUG-2001.
PD
XX
XX 15-FEB-2001; 2001WO-US05075.
PF
XX
XX 18-FEB-2000; 2000US-0507765.
PR
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Klingler TM, Stewart EA, Yue H, Baughn MR;
PI
XX WPI; 2001-522598/57.
XX
XX N-PSDB; AAH74979.
DR
XX
XX A substantially purified GTPase activating protein useful for treating
PT
PT or preventing cell signalling, immune and cell proliferative disorders,
PT
PT including cancer especially colon cancer -
XX
XX Claim 14; Fig 2A-E; 78pp; English.
PS
XX
XX The present sequence represents a human GTPase activating protein,
CC
CC designated GTPAP2. GTPAP polynucleotides and polypeptides are useful
CC
CC for treating or preventing a disease or condition associated with
CC
CC altered expression of GTPase activating proteins, especially cell
CC
CC signalling, immune and cell proliferative disorders, including cancer
CC
CC especially colon cancer. Examples of diseases treatable include
CC
CC arteriosclerosis, diabetes, psoriasis, hepatitis and multiple sclerosis.
CC
CC GTPAP1 polynucleotides are also useful for gene therapy treatments of
CC
CC the diseases.
XX
XX
SQ Sequence 433 AA;
Query Match 30.9%; Score 55; DB 22; Length 433;
Best Local Similarity 52.2%; Pred No. 28;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Caps 0;
OY 2 GLSQOQLVHRSLSQTLRDIOQM 24
DB 224 GLRTEGLFRRSASVTVREIQRL 246

```

Search completed: November 13, 2003, 09:25:20  
Job time : 15.2996 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:19 ; Search time 5.1751 Seconds  
(without alignments)  
286.155 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQQLHRSLSQTLRDQRLPFDEKFTGA 35

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/prodata/1/1aa/5A-COMB pep.\*  
2: /cgn2\_6/prodata/1/1aa/5B-COMB pep.\*  
3: /cgn2\_6/prodata/1/1aa/6A-COMB pep.\*  
4: /cgn2\_6/prodata/1/1aa/6B-COMB pep.\*  
5: /cgn2\_6/prodata/1/1aa/PCFUS-COMB pep.\*  
6: /cgn2\_6/prodata/1/1aa/backfiles1.pap.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	30.9	333	4	US-09-507-765-32
2	55	30.9	433	4	US-09-507-765-30
3	55	30.9	433	4	US-09-507-765-31
4	54.5	30.6	1242	4	US-09-252-991A-25843
5	53	29.8	819	4	US-09-651-656-15
6	53	29.8	819	4	US-09-650-855-15
7	52	29.2	578	4	US-09-252-991A-23906
8	51	28.7	589	4	US-09-252-991A-24834
9	50.5	28.4	354	4	US-09-252-991A-28779
10	50	28.1	221	3	US-09-247-373B-54
11	50	28.1	275	4	US-09-542-749A-2
12	50	28.1	1201	4	US-09-252-991A-32259
13	49	27.5	524	4	US-09-186-276B-56
14	49	27.5	524	4	US-08-842-445-56
15	49	27.5	524	4	US-09-186-188B-56
16	49	27.5	1027	4	US-09-252-991A-23210
17	48.5	27.2	1872	1	US-08-188-582-14
18	48.5	27.2	1893	1	US-08-846-715-14
19	48.5	27.2	1893	1	US-08-188-582-11
20	48.5	27.2	1893	1	US-08-846-715-11
21	48	27.0	286	4	US-09-107-532A-5226
22	48	27.0	1097	4	US-09-252-991A-22579
23	48	27.0	1315	4	US-09-252-991A-22746
24	47.5	26.7	1646	4	US-09-252-991A-29240
25	47	26.4	1000	4	US-09-252-991A-31361
26	47	26.4	10182	4	US-09-134-001C-3159
27	46.5	26.1	65	1	US-08-227-536-3

28 46.5 26.1 65 5 PCT-US95-04682-3  
29 46.5 26.1 885 4 US-09-107-532A-5104  
30 46.5 26.1 935 4 US-09-107-532A-3753  
31 46 25.8 178 4 US-09-252-991A-18906  
32 46 25.8 180 4 US-08-252-991A-31012  
33 46 25.8 482 4 US-08-252-991A-23336  
34 46 25.8 483 4 US-08-238-323-6561  
35 46 25.8 358 4 US-08-252-991A-18908  
36 46 25.8 2101 1 US-08-466-300-4  
37 46 25.8 2101 1 US-08-470-990-4  
38 46 25.8 2101 1 US-08-467-781-4  
39 46 25.8 2101 1 US-08-195-487-4  
40 46 25.8 2101 2 US-08-483-924-4  
41 46 25.8 2101 3 US-09-452-294-1  
42 46 25.8 2101 5 PCT-US93-06160-4  
43 45.5 25.6 118 4 US-08-858-207A-395  
44 45.5 25.6 832 4 US-09-252-991A-19252  
45 45 25.3 52 2 US-08-690-011A-34

## ALIGNMENTS

RESULT 1  
US-09-507-765-32  
Sequence 32, Application US/09507765  
Patent No. 6509155  
GENERAL INFORMATION:  
APPLICANT: Klinger, Tod M.  
APPLICANT: Stewart, Elizabeth A.  
APPLICANT: Yue, Henry  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
FILE REFERENCE: PC-0010 US  
CURRENT APPLICATION NUMBER: US/09/507,765  
CURRENT FILING DATE: 2000-02-18  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PERL Program  
SEQ ID NO 32  
LENGTH: 333  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURES:  
NAME/KEY: misc feature  
OTHER INFORMATION: Incyte ID No. 6509155 96572185  
US-09-507-765-32

Query Match 30.9%; Score 55; DB 4; Length 333;  
Best Local Similarity 52.2%; Pred. No. 6.9;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GLSQQLHRSLSQTLRDQRLM 24  
Db 124 GLTEGLFRSSASVQTVREIQRL 146

RESULT 2  
US-09-507-765-30  
Sequence 30, Application US/09507765  
Patent No. 6509155  
GENERAL INFORMATION:  
APPLICANT: Klinger, Tod M.  
APPLICANT: Stewart, Elizabeth A.  
APPLICANT: Yue, Henry  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
FILE REFERENCE: PC-0010 US  
CURRENT APPLICATION NUMBER: US/09/507,765  
CURRENT FILING DATE: 2000-02-18  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PERL Program  
SEQ ID NO 30  
LENGTH: 433

100

```

; Sequence 23906, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23906
; LENGTH: 578
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23906

Query Match      29.2%; Score 52; DB 4; Length 578;
Best Local Similarity 84.6%; Pred. No. 32;
Matches 11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      6 EOLEHRSLSQTL 18
DB      339 ECLHRSLSRLT 351

RESULT 8
US-09-252-991A-24834
; Sequence 24834, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24834
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24834

Query Match      28.7%; Score 51; DB 4; Length 589;
Best Local Similarity 37.0%; Pred. No. 44;
Matches 10; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY      5 QOELEHRSLSQTLQIQLMLFPDEKE 31
DB      156 EQQNHRRSQLYRQALGFGEADAE 182

RESULT 9
US-09-252-991A-28779
; Sequence 28779, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

```

```

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28779
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28779

Query Match      28.4%; Score 50.5; DB 4; Length 354;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 14; Conservative 6; Mismatches 8; Indels 7; Gaps 2;

QY      2 GLSQE-----QLEHRSLSQTLQIQLMLFPDEKE 31
DB      252 GLAQRLLLPGAPHRPRTLSAVDRQR--FDLRQ 284

RESULT 10
US-09-247-373B-54
; Sequence 54, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247.373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 54
; LENGTH: 221
; TYPE: PRT
; ORGANISM: SOYBEAN
US-09-247-373B-54

Query Match      28.1%; Score 50; DB 3; Length 221;
Best Local Similarity 36.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 8; Indels 4; Gaps 1;

QY      9 EHRERSLQT-----LNDIQRLMFPDEKEFTG 34
DB      116 EEREKSIEKIWEHLVVENOCFGDQKPFPG 145

RESULT 11
US-09-542-749A-2
; Sequence 2, Application US/09542749A
; Patent No. 6428981
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Steen T.
; APPLICANT: Kristensen, Christina L.
; APPLICANT: Kristensen, Tina
; TITLE OF INVENTION: A Bacillus Protein Production Cell
; FILE REFERENCE: 5861.200-US
; CURRENT APPLICATION NUMBER: US/09/542.749A
; CURRENT FILING DATE: 2000-04-04
; PRIOR APPLICATION NUMBER: 60/130,194
; PRIOR FILING DATE: 1999-04-20
; PRIOR APPLICATION NUMBER: PA 1999 00506
; PRIOR FILING DATE: 1999-04-16
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-542-749A-2

Query Match      28.1%; Score 50; DB 4; Length 275;

```

Best Local Similarity 37.0%; Pred. No. 26;  
Matches 10; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

QY 7 OLEHRSLSQTLRDIOQMFLPFDEKFTG 33  
DB 91 ELQGRKAGMQFLRNWQESLFVSKNIT 117

RESULT 12  
US-09-252-991A-32259  
; Sequence 32259, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 32259  
; LENGTH: 1201  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-32259

Query Match 28.1%; Score 50; DB 4; Length 1201;  
Best Local Similarity 39.1%; Pred. No. 1.3e+02;  
Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 3 LSCQLEHRSLSQTLRDIOQMFL 25  
DB 439 VQSRIGHLQSLERLQDRRL 461

RESULT 13  
US-09-186-276B-56  
; Sequence 56, Application US/09186276B  
; Patent No. 6388173  
; GENERAL INFORMATION:  
; APPLICANT: Benfey, Philip  
; APPLICANT: Dilaurenzio, Laura  
; APPLICANT: Wysocka-Diller, Joanna  
; APPLICANT: Malamy, Jocelyn E.  
; APPLICANT: Pysch, Leonard  
; APPLICANT: Helariutta, Yrjo  
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses Thereof  
; FILE REFERENCE: 5914-075-999  
; CURRENT APPLICATION NUMBER: US/09/186,276B  
; CURRENT FILING DATE: 1998-11-05  
; PRIOR APPLICATION NUMBER: 08/842,445  
; PRIOR FILING DATE: 1997-04-24  
; PRIOR APPLICATION NUMBER: 08/638,617  
; PRIOR FILING DATE: 1996-04-26  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 56  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(524)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-186-276B-56

Query Match 27.5%; Score 49; DB 4; Length 524;  
Best Local Similarity 30.0%; Pred. No. 73;  
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 5 OEQLEHRSLSQTLRDIOQMFLPFDEKFTG 34  
DB 66 QNRVHSENMNLSRLEKQLLDDDDSGG 95

RESULT 14  
US-08-842-445-56  
; Sequence 56, Application US/08842445A  
; Patent No. 6441270  
; GENERAL INFORMATION:  
; APPLICANT: Benfey et al.  
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses  
; FILE REFERENCE: 5914-056-999  
; CURRENT APPLICATION NUMBER: US/08/842,445A  
; CURRENT FILING DATE: 1997-04-24  
; EARLIER APPLICATION NUMBER: 08/638,617  
; EARLIER FILING DATE: 1996-04-26  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 56  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Plant  
US-08-842-445-56

Query Match 27.5%; Score 49; DB 4; Length 524;  
Best Local Similarity 30.0%; Pred. No. 73;  
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 5 OEQLEHRSLSQTLRDIOQMFLPFDEKFTG 34  
DB 66 QNRVHSENMNLSRLEKQLLDDDDSGG 95

RESULT 15  
US-09-186-188B-56  
; Sequence 56, Application US/09186188B  
; Patent No. 6455672  
; GENERAL INFORMATION:  
; APPLICANT: Benfey et al.  
; TITLE OF INVENTION: Scarecrow Gene, Promoter and Uses  
; FILE REFERENCE: 5914-074-999  
; CURRENT APPLICATION NUMBER: US/09/186,188B  
; CURRENT FILING DATE: 1998-11-05  
; PRIOR APPLICATION NUMBER: 08/842,445  
; PRIOR FILING DATE: 1997-04-24  
; PRIOR APPLICATION NUMBER: 08/638,617  
; PRIOR FILING DATE: 1996-04-26  
; NUMBER OF SEQ ID NOS: 79  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 56  
; LENGTH: 524  
; TYPE: PRT  
; ORGANISM: Plant  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (1)...(524)  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-09-186-188B-56

Query Match 27.5%; Score 49; DB 4; Length 524;  
Best Local Similarity 30.0%; Pred. No. 73;  
Matches 9; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 5 OEQLEHRSLSQTLRDIOQMFLPFDEKFTG 34  
DB 66 QNRVHSENMNLSRLEKQLLDDDDSGG 95

Search completed: November 13, 2003, 09:28:24

Thu Nov 13 10:41:03 2003

us-09-915-543-15\_copy\_349\_383.ra1

Page 5

Job time : 6.1751 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:27:44 ; Search time 9.39689 Seconds  
(without alignments)  
679.968 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQEQLHRSRSLQTLRDQRMFLPFDEKFTCA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	35	15	US-10-322-579-5
2	178	100.0	1426	15	US-10-322-579-15
3	107	60.1	1115	15	US-10-322-579-17
4	66	37.1	35	15	US-10-322-579-4
5	57.5	32.3	603	15	US-10-128-714-3511
6	57.5	32.3	618	15	US-10-128-714-8511
7	55	30.9	294	12	US-10-053-248-12
8	55	30.9	333	15	US-10-284-753-32
9	55	30.9	290	15	US-10-087-340-21
10	55	30.9	433	15	US-10-284-753-30
11	55	30.9	433	15	US-10-284-753-31
12	55	30.9	464	15	US-10-087-340-14
13	55	30.9	643	15	US-10-087-340-16
14	55	30.9	718	15	US-10-097-340-19
15	55	30.9	751	15	US-10-097-340-18

16	51	28.7	434	9	US-09-815-242-4987
17	51	28.7	448	9	US-09-815-242-10651
18	51	28.7	500	10	US-09-731-872-466
19	51	28.7	500	12	US-09-876-997-466
20	50	28.1	342	12	US-09-244-805-62
21	50	28.1	444	10	US-09-738-826-4566
22	50	28.1	1162	9	US-09-815-242-11828
23	49.5	27.8	464	12	US-10-161-051-119
24	49.5	27.8	804	12	US-10-236-055A-28
25	49.5	27.8	2854	15	US-10-227-610-2
26	49	27.5	201	15	US-10-043-487-323
27	49	27.5	341	12	US-09-244-805-61
28	49	27.5	461	15	US-10-234-432-24
29	49	27.5	524	15	US-10-253-007-56
30	49	27.5	535	15	US-10-043-487-306
31	49	27.5	893	15	US-10-234-432-86
32	49	27.5	1413	12	US-09-840-743-8
33	49	27.5	2000	13	US-10-010-901-29
34	48.5	27.2	898	15	US-10-043-487-277
35	48	27.0	170	9	US-09-916-730-15
36	48	27.0	182	15	US-10-156-761-14821
37	48	27.0	485	12	US-10-287-218-20
38	47.5	26.7	861	9	US-09-815-242-11081
39	47	26.4	529	15	US-10-156-761-10420
40	47	26.4	670	11	US-09-298-523B-63
41	47	26.4	711	11	US-09-298-523B-3
42	47	26.4	810	15	US-10-128-714-8235
43	47	26.4	811	15	US-10-128-714-3235
44	47	26.4	2462	11	US-09-819-104A-5
45	46.5	26.1	100	9	US-09-867-550-2062

## ALIGNMENTS

### RESULT 1

US-10-322-579-5  
; Sequence 5, Application US/10322579  
; Publication NO. US2003014413A1  
; GENERAL INFORMATION:  
; APPLICANT: BASLER, Konrad  
; APPLICANT: BRUNNER, Erich  
; APPLICANT: FROESCH, Barbara  
; APPLICANT: KRAMPS, Thomas  
; APPLICANT: PETER, Oliver  
; TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
; FILE REFERENCE: Q60361  
; CURRENT APPLICATION NUMBER: US/10322,579  
; CURRENT FILING DATE: 2002-12-19  
; PRIOR APPLICATION NUMBER: US/09/915,543  
; PRIOR FILING DATE: 2001-07-27  
; PRIOR APPLICATION NUMBER: 60/221,502  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Human Igs/bcl19  
US-10-322-579-5

Query Match 100.0%; Score 178; DB 15; Length 35;  
Best Local Similarity 100.0%; Pred No. 5e-17; 0; Indels 0; Gaps 0;

Matches 35; Conservative

QY 1 DGLSQEQLHRSRSLQTLRDQRMFLPFDEKFTCA 35

Db 1 DGLSQEQLHRSRSLQTLRDQRMFLPFDEKFTCA 35

### RESULT 2

US-10-322-579-15

Sequence 15, Application US/10322579  
Publication No. US20030114413A1

## GENERAL INFORMATION:

APPLICANT: BASLER, Konrad  
APPLICANT: BRUNNER, Erich  
APPLICANT: FROESCH, Barbara  
APPLICANT: KRAMPS, Thomas

APPLICANT: PETER, Oliver

TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON

FILE REFERENCE: Q60361

CURRENT APPLICATION NUMBER: US/10/322,579

CURRENT FILING DATE: 2002-12-19

PRIOR APPLICATION NUMBER: US/09/915,543

PRIOR FILING DATE: 2001-07-27

PRIOR APPLICATION NUMBER: 60/221,502

PRIOR FILING DATE: 2000-07-28

NUMBER OF SEQ ID NOS: 22

SOFTWARE: PatentIn version 3.1

SEQ ID NO 15

LENGTH: 1426

TYPE: PRT

ORGANISM: Human lgs/bcl9

US-10-322-579-15

Query Match 100.0%; Score 178; DB 15; Length 1426;

Best Local Similarity 100.0%; Pred. No. 3,4e-15;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGLSQQLHRSRSLQTLRDIOQLMPLPDEKFTGA 35

DB 349 DGLSQQLHRSRSLQTLRDIOQLMPLPDEKFTGA 383

## RESULT 3

US-10-322-579-17

Sequence 17, Application US/10322579

Publication No. US20030114413A1

## GENERAL INFORMATION:

APPLICANT: BASLER, Konrad

APPLICANT: BRUNNER, Erich

APPLICANT: FROESCH, Barbara

APPLICANT: KRAMPS, Thomas

APPLICANT: PETER, Oliver

TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON

FILE REFERENCE: Q60361

CURRENT APPLICATION NUMBER: US/10/322,579

CURRENT FILING DATE: 2002-12-19

PRIOR APPLICATION NUMBER: US/09/915,543

PRIOR FILING DATE: 2001-07-27

PRIOR APPLICATION NUMBER: 60/221,502

PRIOR FILING DATE: 2000-07-28

NUMBER OF SEQ ID NOS: 22

SOFTWARE: PatentIn version 3.1

SEQ ID NO 17

LENGTH: 1115

TYPE: PRT

ORGANISM: Human lgs-1

US-10-322-579-17

Query Match 60.1%; Score 107; DB 15; Length 1115;

Best Local Similarity 87.5%; Pred. No. 8,9e-06;

Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 GLSQQLHRSRSLQTLRDIOQLM 25

DB 76 GLSQQLHRSRSLQTLRDIOQLM 99

## RESULT 4

US-10-322-579-4

Sequence 4, Application US/10322579

Publication No. US20030114413A1

## GENERAL INFORMATION:

APPLICANT: BASLER, Konrad

APPLICANT: BRUNNER, Erich

APPLICANT: FROESCH, Barbara

APPLICANT: KRAMPS, Thomas

APPLICANT: PETER, Oliver

TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON

FILE REFERENCE: Q60361

CURRENT APPLICATION NUMBER: US/10/322,579

CURRENT FILING DATE: 2002-12-19

PRIOR APPLICATION NUMBER: US/09/915,543

PRIOR FILING DATE: 2001-07-27

PRIOR APPLICATION NUMBER: 60/221,502

PRIOR FILING DATE: 2000-07-28

NUMBER OF SEQ ID NOS: 22

SOFTWARE: PatentIn version 3.1

SEQ ID NO 4

LENGTH: 35

TYPE: PRT

ORGANISM: Drosophila lgs

US-10-322-579-4

Query Match 37.1%; Score 66; DB 15; Length 35;

Best Local Similarity 31.4%; Pred. No. 0.054;

Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 1 DGLSQQLHRSRSLQTLRDIOQLMPLPDEKFTGA 35

DB 1 ENLTPQQRHREQLAKIKKXNQFLPFENSVGA 35

## RESULT 5

US-10-128-714-3511

Sequence 3511, Application US/10128714

Publication No. US20030119013A1

## GENERAL INFORMATION:

APPLICANT: Jiang, Bo

APPLICANT: Hu, Wengqi

APPLICANT: Tishkoff, Daniel

APPLICANT: Zamudio, Carlos

APPLICANT: Broshkin, Alexey M

APPLICANT: Lemieux, Sebastien M

TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and

TITLE OF INVENTION: Methods of Use

FILE REFERENCE: 10182-018-999

CURRENT APPLICATION NUMBER: US/10/128,714

CURRENT FILING DATE: 2002-04-23

PRIOR APPLICATION NUMBER: US 60/285,697

PRIOR FILING DATE: 2001-04-23

PRIOR APPLICATION NUMBER: US 60/287,066

PRIOR FILING DATE: 2001-04-27

PRIOR APPLICATION NUMBER: US 60/295,890

PRIOR FILING DATE: 2001-06-05

PRIOR APPLICATION NUMBER: US 60/303,899

PRIOR FILING DATE: 2001-07-09

PRIOR APPLICATION NUMBER: US 60/316,362

PRIOR FILING DATE: 2001-08-31

NUMBER OF SEQ ID NOS: 8603

SOFTWARE: PatentIn version 3.1

SEQ ID NO 3511

LENGTH: 603

TYPE: PRT

ORGANISM: Aspergillus fumigatus

US-10-128-714-3511

Query Match 32.3%; Score 57.5; DB 15; Length 603;

Best Local Similarity 31.4%; Pred. No. 19;

Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;

QY 1 DGLSQQLHRSRSLQTLRDIOQLMPLPDEKFTGA 35

DB 1 DGLSQQLHRSRSLQTLRDIOQLMPLPDEKFTGA 35

45 DGVETEKIREKD---EVEKKLRMLFGDDGFVGA 76

Db

## RESULT 6

US-10-128-714-8511  
; Sequence 5513, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Meng, Daniel  
; APPLICANT: Tishkoff, Carlos  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroskin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
; FILE REFERENCE: 10182-018-999  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 8511  
; TYPE: PRT  
; LENGTH: 618  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-8511

Query Match 32.3%; Score 57.5; DB 15; Length 618;

Best Local Similarity 31.4%; Pred. No. 20;  
Matches 11; Conservative 11; Mismatches 10; Indels 3; Gaps 1;

QY 1 DLSQQLHRRSLQTLRDQRMFLPDPKXFTGA 35

Db 45 DGVETEKIREKD---EVEKKLRMLFGDDGFVGA 76

## RESULT 7

US-10-053-248-12  
; Sequence 12, Application US/1005248  
; Publication No. US2003014188A1  
; GENERAL INFORMATION:  
; APPLICANT: Lin, BiaoYang  
; TITLE OF INVENTION: Androgen Regulated Nucleic Acid  
; FILE REFERENCE: P-15 4814  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR FILING DATE: 2002-01-15  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; TYPE: PRT  
; LENGTH: 294  
; ORGANISM: Homo sapiens  
US-10-053-248-12

Query Match 30.9%; Score 55; DB 12; Length 294;

Best Local Similarity 52.2%; Pred. No. 19;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDQRM 24

Db 224 GLRTEGLFRSASVQTVREIQL 246

## RESULT 8

US-10-284-753-32  
; Sequence 32, Application US/10284753  
; Publication No. US20030129655A1  
; GENERAL INFORMATION:  
; APPLICANT: Klinger, Tod M.  
; APPLICANT: Stewart, Elizabeth A.  
; APPLICANT: Yue, Henry  
; APPLICANT: Baughn, Mawiah R.  
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
; FILE REFERENCE: PC-0010-1 CIP  
; CURRENT FILING DATE: 2002-10-29  
; PRIOR FILING DATE: 2002-10-29  
; PRIOR APPLICATION NUMBER: 09/507,765  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PERL Program  
; SEQ ID NO 32  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. US20030129655A1 96572185  
US-10-284-753-32

Query Match 30.9%; Score 55; DB 16; Length 333;

Best Local Similarity 52.2%; Pred. No. 21;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDQRM 24

Db 124 GLRTEGLFRSASVQTVREIQL 146

## RESULT 9

US-10-097-340-21  
; Sequence 21, Application US/10097340  
; Publication No. US20030087250A1  
; GENERAL INFORMATION:  
; APPLICANT: John Monahan  
; APPLICANT: Manjula GANNAVARAPU  
; APPLICANT: Sebastian HOERSCH  
; APPLICANT: Shubhangi KAWATKAR  
; APPLICANT: Steve G. KOVATS  
; APPLICANT: Rachel E. MEYERS  
; APPLICANT: Michael MORRISSEY  
; APPLICANT: Peter OLANDT  
; APPLICANT: Ami SEN  
; APPLICANT: Peter VEIBY  
; APPLICANT: Gordon B. MILLS  
; APPLICANT: Robert C. BAST, Jr.  
; APPLICANT: Karen LU  
; APPLICANT: Rosemarie SCHMANDT  
; APPLICANT: Xumei ZHAO  
; APPLICANT: Karen GLATT  
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
; FILE REFERENCE: MRI-030  
; CURRENT FILING DATE: 2002-03-14  
; PRIOR FILING DATE: 2002-03-14  
; PRIOR APPLICATION NUMBER: 60/276,025  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/325,149  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/276,026  
; PRIOR FILING DATE: 2001-03-14  
; PRIOR APPLICATION NUMBER: 60/324,967  
; PRIOR FILING DATE: 2001/09/26  
; PRIOR APPLICATION NUMBER: 60/311,732  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/325,102  
US-10-097-340-21

/ PRIOR FILING DATE: 2001-09-26  
/ PRIOR APPLICATION NUMBER: 60/323,580  
/ PRIOR FILING DATE: 2001-09-19  
/ NUMBER OF SEQ ID NOS: 363  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 21  
/ LENGTH: 390  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-097-340-21

Query Match 30.9%; Score 55; DB 15; Length 390;  
Best Local Similarity 52.2%; Pred. No. 26;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDQRM 24  
DB 181 GLRTEGLFRSSASVQTVREIQL 203

RESULT 10  
US-10-284-753-30  
/ Sequence 30, Application US/10284753  
/ Publication No. US20030129655A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Klingert, Tod M.  
/ APPLICANT: Stewart, Elizabeth A.  
/ APPLICANT: Yue, Henry  
/ APPLICANT: Baughn, Mariah R.  
/ TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
/ FILE REFERENCE: PC-0010-1 CIP  
/ CURRENT APPLICATION NUMBER: US/10/284,753  
/ PRIOR FILING DATE: 2002-10-23  
/ PRIOR APPLICATION NUMBER: 09/507,765  
/ PRIOR FILING DATE: 2000-02-18  
/ NUMBER OF SEQ ID NOS: 33  
/ SOFTWARE: PERL Program  
/ SEQ ID NO 30  
/ LENGTH: 433  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ NAME/KEY: misc feature  
/ OTHER INFORMATION: Incyte ID No. US20030129655A1 3068538CD1  
US-10-284-753-30

Query Match 30.9%; Score 55; DB 16; Length 433;  
Best Local Similarity 52.2%; Pred. No. 29;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDQRM 24  
DB 224 GLRTEGLFRSSASVQTVREIQL 246

RESULT 11  
US-10-284-753-31  
/ Sequence 31, Application US/10284753  
/ Publication No. US20030129655A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Klingert, Tod M.  
/ APPLICANT: Stewart, Elizabeth A.  
/ APPLICANT: Yue, Henry  
/ APPLICANT: Baughn, Mariah R.  
/ TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS  
/ FILE REFERENCE: PC-0010-1 CIP  
/ CURRENT APPLICATION NUMBER: US/10/284,753  
/ PRIOR FILING DATE: 2002-10-23  
/ PRIOR APPLICATION NUMBER: 09/507,765  
/ PRIOR FILING DATE: 2000-02-18  
/ NUMBER OF SEQ ID NOS: 33  
/ SOFTWARE: PERL Program  
/ SEQ ID NO 31

/ LENGTH: 433  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
/ NAME/KEY: misc feature  
/ OTHER INFORMATION: Incyte ID No. US20030129655A1 404424.5.pseq  
US-10-284-753-31

Query Match 30.9%; Score 55; DB 16; Length 433;  
Best Local Similarity 52.2%; Pred. No. 29;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDQRM 24  
DB 224 GLRTEGLFRSSASVQTVREIQL 246

RESULT 12  
US-10-097-340-14  
/ Sequence 14, Application US/10097340  
/ Publication No. US20030087250A1  
/ GENERAL INFORMATION:  
/ APPLICANT: John MONAHAN  
/ APPLICANT: Marjulla GANNAVAREPU  
/ APPLICANT: Sebastian HOERSCH  
/ APPLICANT: Shubhangi KAMATKAR  
/ APPLICANT: Steve G. KOVATS  
/ APPLICANT: Rachel E. MEYERS  
/ APPLICANT: Michael MORRISSEY  
/ APPLICANT: Peter OLANDT  
/ APPLICANT: Ami SEN  
/ APPLICANT: Peter VEIBY  
/ APPLICANT: Gordon B. MILLS  
/ APPLICANT: Robert C. EAST, Jr.  
/ APPLICANT: Karen LU  
/ APPLICANT: Rosemarie SCHMANDT  
/ APPLICANT: Xumei ZHAO  
/ APPLICANT: Karen ZHANG

/ TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
/ Assessment, Prevention, and Therapy of Ovarian Cancer  
/ FILE REFERENCE: MRI-030  
/ CURRENT APPLICATION NUMBER: US/10/097,340  
/ PRIOR FILING DATE: 2002-03-14  
/ PRIOR APPLICATION NUMBER: 60/276,025  
/ PRIOR FILING DATE: 2001-03-14  
/ PRIOR APPLICATION NUMBER: 60/325,149  
/ PRIOR FILING DATE: 2001-09-26  
/ PRIOR APPLICATION NUMBER: 60/276,026  
/ PRIOR FILING DATE: 2001-03-14  
/ PRIOR APPLICATION NUMBER: 60/324,967  
/ PRIOR FILING DATE: 2001/09/26  
/ PRIOR APPLICATION NUMBER: 60/311,732  
/ PRIOR FILING DATE: 2001-08-10  
/ PRIOR APPLICATION NUMBER: 60/325,102  
/ PRIOR FILING DATE: 2001-09-26  
/ PRIOR APPLICATION NUMBER: 60/323,580  
/ NUMBER OF SEQ ID NOS: 363  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 14  
/ LENGTH: 464  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-097-340-14

Query Match 30.9%; Score 55; DB 15; Length 464;  
Best Local Similarity 52.2%; Pred. No. 31;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 2 GLSQQLHRRSLQTLRDQRM 24  
DB 255 GLRTEGLFRSSASVQTVREIQL 277

## RESULT 13

US-10-097-340-16

; Sequence 16, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN

; APPLICANT: Manjula GHANAVARAPU

; APPLICANT: Sebastian HOERSCH

; APPLICANT: Shubhangi KAWATKAR

; APPLICANT: Steve G. KOVATS

; APPLICANT: Rachel E. MEYERS

; APPLICANT: Michael MORRISSEY

; APPLICANT: Peter OLANDT

; APPLICANT: Ami SEN

; APPLICANT: Peter VEIBY

; APPLICANT: Gordon B. MILLS

; APPLICANT: Robert C. BAST, Jr.

; APPLICANT: Karen LU

; APPLICANT: Rosemarie SCHMANDT

; APPLICANT: Xumel ZHAO

; APPLICANT: Karen GLATT

; APPLICANT: Karen GLATT

; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,

; FILE REFERENCE: MRI-030

; CURRENT APPLICATION NUMBER: US/10/097,340

; CURRENT FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: 60/276,025

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/325,149

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/276,026

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/324,967

; PRIOR FILING DATE: 2001/09/26

; PRIOR APPLICATION NUMBER: 60/311,732

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: 60/325,102

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/323,580

; PRIOR FILING DATE: 2001-09-19

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 16

; LENGTH: 643

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-097-340-16

Query Match 30.9%; Score 55; DB 15; Length 643;

Best Local Similarity 52.2%; Pred. No. 45;

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy

2 GLSQQLERHSLSQTLRDQIRM 24

Db

434 GLTEGLFRSASVQTVREIQRL 456

## RESULT 14

US-10-097-340-19

; Sequence 19, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN

; APPLICANT: Manjula GHANAVARAPU

; APPLICANT: Sebastian HOERSCH

; APPLICANT: Shubhangi KAWATKAR

; APPLICANT: Steve G. KOVATS

; APPLICANT: Rachel E. MEYERS

; APPLICANT: Michael MORRISSEY

; APPLICANT: Peter OLANDT

; APPLICANT: Ami SEN

; APPLICANT: Peter VEIBY

; APPLICANT: Gordon B. MILLS  
 ; APPLICANT: Robert C. BAST, Jr.  
 ; APPLICANT: Karen LU  
 ; APPLICANT: Rosemarie SCHMANDT  
 ; APPLICANT: Xumel ZHAO  
 ; APPLICANT: Karen GLATT  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
 ; FILE REFERENCE: MRI-030  
 ; CURRENT APPLICATION NUMBER: US/10/097,340  
 ; CURRENT FILING DATE: 2002-03-14  
 ; PRIOR APPLICATION NUMBER: 60/276,025  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/325,149  
 ; PRIOR FILING DATE: 2001-09-26  
 ; PRIOR APPLICATION NUMBER: 60/276,026  
 ; PRIOR FILING DATE: 2001-03-14  
 ; PRIOR APPLICATION NUMBER: 60/324,967  
 ; PRIOR FILING DATE: 2001/09/26  
 ; PRIOR APPLICATION NUMBER: 60/311,732  
 ; PRIOR FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: 60/325,102  
 ; PRIOR FILING DATE: 2001-09-26  
 ; PRIOR APPLICATION NUMBER: 60/323,580  
 ; PRIOR FILING DATE: 2001-09-19  
 ; NUMBER OF SEQ ID NOS: 363  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 19  
 ; LENGTH: 718  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-097-340-19

Query Match 30.9%; Score 55; DB 15; Length 718;

Best Local Similarity 52.2%; Pred. No. 52;

Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy

2 GLSQQLERHSLSQTLRDQIRM 24

Db

509 GLTEGLFRSASVQTVREIQRL 531

## RESULT 15

US-10-097-340-18

; Sequence 18, Application US/10097340

; Publication No. US20030087250A1

; GENERAL INFORMATION:

; APPLICANT: John MONAHAN

; APPLICANT: Manjula GHANAVARAPU

; APPLICANT: Sebastian HOERSCH

; APPLICANT: Shubhangi KAWATKAR

; APPLICANT: Steve G. KOVATS

; APPLICANT: Rachel E. MEYERS

; APPLICANT: Michael MORRISSEY

; APPLICANT: Peter OLANDT

; APPLICANT: Ami SEN

; APPLICANT: Peter VEIBY

; APPLICANT: Gordon B. MILLS

; APPLICANT: Robert C. BAST, Jr.

; APPLICANT: Karen LU

; APPLICANT: Rosemarie SCHMANDT

; APPLICANT: Xumel ZHAO

; APPLICANT: Karen GLATT

; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,

; FILE REFERENCE: MRI-030

; CURRENT APPLICATION NUMBER: US/10/097,340

; CURRENT FILING DATE: 2002-03-14

; PRIOR APPLICATION NUMBER: 60/276,025

; PRIOR FILING DATE: 2001-03-14

; PRIOR APPLICATION NUMBER: 60/325,149

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: 60/276,026

```

/ PRIOR FILING DATE: 2001-09-14
/ PRIOR APPLICATION NUMBER: 60/324,967
/ PRIOR FILING DATE: 2001/09/36
/ PRIOR APPLICATION NUMBER: 60/311,732
/ PRIOR FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: 60/325,102
/ PRIOR FILING DATE: 2001-09-26
/ PRIOR APPLICATION NUMBER: 60/323,580
/ PRIOR FILING DATE: 2001-09-19
/ NUMBER OF SEQ ID NOS: 363
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 18
/ LENGTH: 751
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: 1...751
/ OTHER INFORMATION: Xaa = Any Amino Acid .
US-00-997-340-18

```

Query Match 30.9%; Score 55; DB 15; Length 751;  
Best Local Similarity 52.2%; Pred. No. 54;  
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Search completed: November 13, 2003, 09:39:45  
Job time : 9.39689 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 4.76654 Seconds  
(without alignments)  
706.153 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQQLHRESLQLTRDIQRMFLPFDEKFTGA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168602 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 10%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 76.4

2: PIR1\*

3: PIR2\*

4: PIR3\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	32.0	1034	2 AB0551	exonuclease SbcC [
2	56	31.5	584	2 S51882	topoisomerase I-re
3	56	31.5	859	2 T46372	hypothetical prote
4	55	30.9	287	2 A82895	transcription regu
5	55	30.9	295	2 H97670	hex regulon repres
6	55	30.9	643	2 B59436	Rho GTPase activat
7	54.5	30.6	1171	2 S14066	phycochrome B - ri
8	53.5	30.1	4957	2 T03455	ALR protein - huma
9	53.5	30.1	5262	2 T03454	hypothetical prote
10	53	29.8	227	2 B70438	hypothetical prote
11	53	29.8	376	1 B69957	gamma-D-glutamyl-L
12	53	29.8	818	2 S62790	mismatch DNA recog
13	52.5	29.5	242	2 B70366	hypothetical prote
14	52.5	29.5	705	2 T24343	hypothetical prote
15	52.5	29.5	1039	2 T14802	phycochrome B - so
16	52	29.2	332	2 B47017	probable transcrip
17	52	29.2	332	2 D82581	transcription init
18	52	29.2	572	2 AD2541	pyruvate dehydroge
19	52	29.2	1009	2 S61174	hypothetical prote
20	51	28.7	102	2 AH0216	conserved hypoteth
21	51	28.7	237	2 A49940	probable alpha hel
22	51	28.7	237	2 E91056	probable alpha hel
23	51	28.7	237	2 A85901	probable alpha hel
24	51	28.7	329	2 D96834	hypothetical prote
25	51	28.7	477	2 T18801	hypothetical prote
26	51	28.7	899	1 GNMVNM	pol polyprotein -
27	51	28.7	1047	2 G90684	ATP-dependent deBN
28	51	28.7	1047	2 C85535	development protei
29	51	28.7	1464	2 S58984	

## ALIGNMENTS

### RESULT 1

AB0551

exonuclease SbcC [imported] - Salmonella enterica subsp. enterica serovar Typhi (stra C:Species: Salmonella enterica subsp. enterica serovar Typhi  
A:Note: this species has also been called Salmonella typhi  
C:Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
A:Accession: AB0551

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church-  
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr  
S.; Moulie, S.; O'Garra, P.  
Nature 413, 848-852, 2001

A:Authors: Barry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,  
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0551; MUID:21534947; PMID:11677608

A:Accession: AB0551

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1034 <P&S>

A:Cross-references: GB:AL513382; PIDN:CAD08850.1; PID:gl6501663; GSPDB:GN00176

C:Genetics:

A:Gene: ST10429

C:Superfamily: sbcC protein

Query Match 32.0%; Score S7; DB 2; Length 1034;

Best Local Similarity 42.9%; Pred. No. 26;

Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 3 LSQQLHRESLQLTRDIQRMFLPFDEK 30

DB 213 LADSQQLHRESLQLTRDIQRMFLPFDEK 240

RESULT 2

S51882

topoisomerase I-related protein TRF4 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein HRC584; protein O0716; protein YOL115W

C:Species: Saccharomyces cerevisiae

C:Date: 05-May-1995 #sequence\_revision 03-Aug-1995 #text\_change 21-Jul-2000

A:Accession: S51882; S51858; S58774; S66811

R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.

submitted to the EMBL Data Library, January 1995

A:Description: Sequence analysis of a 44kb DNA fragment of yeast chromosome XV includi

and a Delta.

A:Reference number: S51848

A:Accession: S51882

A:Molecule type: DNA

A:Residues: 1-584 <VAN>

A:Cross-references: EMBL:248149; NID:g63234; PID:g63237

R:Vandenbol, M.; Durand, P.; Portetelle, D.; Hilger, F.

Yeast 11, 1069-1075, 1995

A:Title: Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV including th

a delta element.

A;reference number: ES9436  
A;Accession: J33438  
A;Species: *Saccharomyces cerevisiae*  
A;Molecule type: mRNA  
A;Residues: 1-643 <GOW>  
A;cross-references: GB:CAB90249; PID:g771l011; PIDN:CAB90248.1  
  
Query Match            30.9%; Score 55; DB 2; Length 643;  
Best Local Similarity   52.2%; Pred. No. 28;



Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 QY 2 GLSQEQLHRSRSLQTLRIQM 24  
 Db 434 GLRTEGLFRASQVTVREIQL 456

## RESULT 7

S14065  
 phytochrome B - rice  
 C:Species: Oryza sativa (rice)  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 21-Jul-2000  
 C:Accession: S14065  
 R:Dehesh, K.; Tepperman, J.; Christensen, A.H.; Quail, P.H.  
 Mol. Gen. Genet. 225, 305-313, 1991  
 A:Title: phyB is evolutionarily conserved and constitutively expressed in rice seedling  
 A:Reference number: S14065; MUID:91172131; PMID:2005872  
 A:Accession: S14065  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1171 <DEH>  
 A:Cross-references: GB:X57563; NID:g6469490; PIDN:CAA40795.2; PID:g6469491  
 C:Genetics:  
 A:Gene: phyB  
 C:Superfamily: phytochrome; phytochrome homology  
 C:Keywords: chromoprotein; photoreceptor; phytochromobilin; transcription regulation  
 F:103-623/Domain: phytochrome homology <PHYT>  
 F:364/Binding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 30.6%; Score 54.5; DB 2; Length 1171;  
 Best Local Similarity 48.1%; Pred. No. 63;  
 Matches 13; Conservative 4; Mismatches 5; Indels 5; Gaps 1;  
 QY 3 LSQEQLEHRSRSLQTLRIQM 29  
 Db 1019 VSQVMIQLRDLQLRDI-----PDE 1040

## RESULT 8

T03455  
 ALR protein - human  
 C:Species: Homo sapiens (man)  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
 C:Accession: T03455  
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,  
 Oncogene 15, 549-560, 1997  
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolog  
 A:Reference number: 214954; MUID:97388474; PMID:9247308  
 A:Accession: T03455  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-4957 <PRA>  
 A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AACS1735.1; PID:g2358287  
 C:Genetics:  
 A:Gene: ALR  
 A:Map position: 12  
 C:Superfamily: human ALR protein  
 C:Keywords: alternative splicing

Query Match 30.1%; Score 53.5; DB 2; Length 4957;  
 Best Local Similarity 42.5%; Pred. No. 4.3e+02;  
 Matches 17; Conservative 4; Mismatches 12; Indels 7; Gaps 2;  
 QY 2 GLSQEQLHRSRSLQTLRIQM 35  
 Db 2091 GLSQTELE-KQRQRLRELLIRQIQRNLTLRQKETAA 2129

## RESULT 9

T03454  
 ALR protein - human  
 C:Species: Homo sapiens (man)  
 C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000

C:Accession: T03454  
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano,  
 Oncogene 15, 549-560, 1997  
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homo  
 A:Reference number: 214954; MUID:97388474; PMID:9247308  
 A:Accession: T03454  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-5262 <PRA>  
 A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AACS1734.1; PID:g2358285  
 C:Genetics:  
 A:Gene: ALR  
 A:Map position: 12  
 C:Superfamily: human ALR protein  
 C:Keywords: alternative splicing

Query Match 30.1%; Score 53.5; DB 2; Length 5262;  
 Best Local Similarity 42.5%; Pred. No. 4.5e+02;  
 Matches 17; Conservative 4; Mismatches 12; Indels 7; Gaps 2;  
 QY 2 GLSQEQLHRSRSLQTLRIQM 35  
 Db 2396 GLSQTELE-KQRQRLRELLIRQIQRNLTLRQKETAA 2434

RESULT 10  
 B70438  
 hypothetical protein aq\_1596 - Aquifex aeolicus  
 C:Species: Aquifex aeolicus  
 C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 05-Nov-1999  
 C:Accession: B70438  
 R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;  
 Nature 392, 353-358, 1998  
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
 A:Reference number: A70300; MUID:98196666; PMID:9537320  
 A:Accession: B70438  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-227 <AAQF>  
 A:Cross-references: GB:AE000747; NID:g2983944; PIDN:AAC07503.1; PID:g2983955; GB:AE000  
 A:Experimental source: strain VF5  
 C:Genetics:  
 A:Gene: aq\_1596

Query Match 29.8%; Score 53; DB 2; Length 227;  
 Best Local Similarity 41.4%; Pred. No. 16;  
 Matches 12; Conservative 7; Mismatches 8; Indels 2; Gaps 1;  
 QY 5 QEQLEHRSRSLQTLRIQM 31  
 Db 33 KELLEEREKULETIRSYEEKLDSFEKQ 61

RESULT 11  
 B6957  
 gamma-D-glutamyl-L-diamino acid endopeptid homolog yggt - Bacillus subtilis  
 C:Species: Bacillus subtilis  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
 C:Accession: B6957  
 R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bert  
 C: Bron, S.; Brouillet, S.; Bruchi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; C  
 A: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 ick, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M  
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue  
 Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel  
 Rieger, M.; Rivolta, C.; Rocha, E. Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlo  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Skiguchi, J.; Sokowska, A.; S  
 akuchi, M.; Tamakoshi, A.; Tanaka, R.; Terpetra, P.; Tognoni, A.; Tosato, V.; Uchiyam  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033; PMID:9384377  
 A:Accession: E69957  
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-376 <KUN>  
 A:Cross-references: GB:299116; GB:AL009126; NID:92634723; PIDN:CAB14414.1; PID:g2634917  
 A:Experimental source: strain 168  
 A:Gene: ydgT  
 C:Genetics:  
 C:Superfamily: endopeptidase I

Query Match 29.8%; Score 53; DB 1; Length 376;  
 Best Local Similarity 52.6%; Pred. No. 28;  
 Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 9 EHRRSLQTLRDIQRMFLFP 27  
 DB 50 EHSKELQIQIKRFLQP 68

## RESULT 12

S62790  
 mismatch DNA recognition protein muts [validated] - *Thermus aquaticus* (fragment)  
 C:Species: *Thermus aquaticus*  
 C>Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 26-May-2000  
 C:Accession: S62790

R.Takamatsu, S.; Kato, R.; Kuramitsu, S.

Nucleic Acids Res. 24, 640-647, 1996  
 A:Title: Mismatch DNA recognition protein from an extremely thermophilic bacterium, *Thermus*  
 A:Reference number: S62790; MUID:96177563; PMID:8604304

A:Accession: S62790

A>Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-818 <TAK>

A:Cross-references: EMBL:D63810

A>Note: the source is designated as *Thermus thermophilus*

C:Genetics:

A:Gene: muts

C:Function:

A:Description: implicated in DNA mismatch repair; binds to DNA and specifically recognizes

attached DNA [validated, MUID:96177563]

C:Superfamily: DNA mismatch repair protein muts

Query Match 29.8%; Score 53; DB 2; Length 818;  
 Best Local Similarity 42.3%; Pred. No. 67;  
 Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 6 EQLHRRSLQTLRDIQRMFLFPDEKE 31  
 DB 465 EKVPQEVPRVQLKDRQRYTLPEMKE 490

## RESULT 13

B70366

hypothetical protein aq\_755 - *Aquifex aeolicus*

C:Species: *Aquifex aeolicus*

C>Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 18-Aug-2000

C:Accession: B70366

R.Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'N

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A:Reference number: A70500; MUID:96156666; PMID:9537320

A:Accession: B70366

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-242 <AQF>

A:Cross-references: GB:AE000706; NID:g2983327; PIDN:AAC06923.1; PID:g2983334; GB:AE00069

A:Experimental source: strain VFS

C:Genetics:

A:Gene: aq\_755

C:Superfamily: *Aquifex aeolicus* hypothetical protein aq\_755

Query Match 29.5%; Score 52.5; DB 2; Length 242;  
 Best Local Similarity 45.8%; Pred. No. 20;  
 Matches 11; Conservative 7; Mismatches 5; Indels 1; Gaps 1;

QY 10 HRERSLQTLRDIQRMFLFPDEKEF 32  
 DB 60 HKRTSLRKFRVRIEMVFEAEKEF 83

## RESULT 14

T24343

hypothetical protein T02B5.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 02-Jun-2000

C:Accession: T24343

R.McMurray, A.

submitted to the EMBL Data Library, October 1996

A:Reference number: Z19878

A:Accession: T24343

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-705 <WIL>

A:Cross-references: EMBL:Z81112; PIDN:CAB03272.1; GSPDB:GN00023; CESP:T02B5.1

A:Experimental source: clone T02B5

C:Genetics:

A:Gene: CESP:T02B5.1

A:Map position: 5

A:Introns: 22/2; 88/2; 117/3; 185/1; 221/3; 280/3; 349/2; 427/2; 532/2; 637/1

C:Superfamily: cholinesterase; cholinesterase homology

Query Match 29.5%; Score 52.5; DB 2; Length 705;  
 Best Local Similarity 36.8%; Pred. No. 66;  
 Matches 14; Conservative 7; Mismatches 12; Indels 5; Gaps 1;

QY 3 LSOROLEH-----RERSLQTLRDIQRMFLFPDEKEFTGA 35  
 DB 273 LSEQVNTYSLRKRSQAQILDQAQLLQNLQSTYFLGA 310

## RESULT 15

T14802

phytochrome B - sorghum (fragment)

C:Species: *Sorghum bicolor* (sorghum)

C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 05-May-2000

C:Accession: T14802

R.Childe, K.L.; Miller, P.R.; Cordonnier-Pratt, M.M.; Pratt, L.H.; Morgan, P.W.; Muller

submitted to the EMBL Data Library, April 1996

A:Description: The Sorghum bicolor photoperiod sensitivity gene, Ma3, encodes a phytochrome

A:Reference number: Z18185

A:Accession: T14802

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-1039 <CHI>

A:Cross-references: EMBL:U56730; NID:g1800216; PID:g1800217

A:Experimental source: cultivar 58M

C:Genetics:

A:Gene: PHYB

A:Note: Intron positions not resolved (incomplete sequence)

C:Superfamily: phytochrome; phytochrome homology

C:Keywords: chromophore; photoreceptor; phytochromobilin

F;23/finding site: phytochromobilin (Cys) (covalent) #status predicted

Query Match 29.5%; Score 52.5; DB 2; Length 1039;  
 Best Local Similarity 51.9%; Pred. No. 1e+02;  
 Matches 14; Conservative 2; Mismatches 6; Indels 5; Gaps 1;

QY 3 LSQEQLEHRRSLQTLRDIQRMFLFPDE 29  
 DB 887 VSQAMLLLRERDLQIRDI-----PDE 908

, Thu Nov 13 10:41:04 2003

us-09-915-543-15\_copy\_349\_383.rpr

Page 5

Search completed: November 13, 2003, 09:29:17  
Job time : 7.76654 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 3.1323 Seconds  
(without alignments)  
525.472 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQEQLHRSRLQTLRDQLRMLPDEKEFTGA 35

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	178	100.0	1426	1	000512 homo sapien
2	66	37.1	1469	1	BCU9 HUMAN
3	56	31.5	584	1	Q961d9 drosophila
4	55	30.9	426	1	TRF4 YEAST
5	55	30.9	718	1	HEM1 SYNEL
6	54.5	30.6	174	1	Q9ns90 homo sapien
7	54.5	30.6	1171	1	PHG8 HUMAN
8	53	29.8	376	1	Q94jir7 mus musculus
9	53	29.8	817	1	P25764 oryza sativ
10	53	29.8	818	1	P54497 bacillus su
11	52.5	29.5	180	1	Q94ix6 thermus cal
12	52.5	29.5	190	1	Q56239 thermus the
13	52.5	29.5	242	1	Q9nc33 homo sapien
14	52.5	29.5	1178	1	Q8x529 fusobacteri
15	52	29.2	332	1	O66957 aquifex aeo
16	52	29.2	311	1	P93527 sorghum bic
17	52	29.2	1009	1	Q03065 anabaena sp
18	52	29.2	1009	1	Q56215 thermus agu
19	51	28.7	237	1	Q06407 saccharomy
20	51	28.7	899	1	P37328 escherichia
21	50	28.1	425	1	P03365 mouse mamma
22	49.5	27.8	101	1	Q92794 homo sapien
23	49.5	27.8	310	1	P02642 gallus gall
24	49.5	27.8	310	1	Q90555 bacteroides
25	49	27.5	328	1	Q9K528 bacillus ha
26	49	27.5	661	1	Q9n10 staphylococ
27	49	27.5	1297	1	Q10369 schizosach
28	49	27.5	1297	1	Q95136 arabidopsis
29	48.5	27.2	184	1	Q01570 homo sapien
30	48.5	27.2	1073	1	P45870 bacillus su
31	48.5	27.2	1132	1	CAR6 METWA
32	48.5	27.2	1505	1	P29130 nicotiana t
33	48.5	27.2	1872	1	Q70608 rattus norv

## ALIGNMENTS

### RESULT 1

ID	BCU9 HUMAN	STANDARD	PRT	1426 AA
AC	O00512;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).			
GN	BCU9			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	[1]			
OX	NCEI_TaxID=9606;			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Fetal brain;			
RX	MEDLINE=98150621; PubMed=9490669;			
RA	Willis T.G., Zaiberg I.R., Coignet L.J.A., Wlodarska J., Stul M.,			
RA	Jadavai D.W., Bastard C., Freleaux J.G., Catovsky D., Silva M.H.M.,			
RA	Dyer M.J.S.;			
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel			
RT	gene (BCU9) at chromosome 1q21.";			
RL	Blood 91:1873-1881(1998).			
RN	[2]			
RN	FUNCTION.			
RX	MEDLINE=21952490; PubMed=1955446;			
RA	Krams T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,			
RA	Murphy M., Ziegler S., Basler K.;			
RT	"Wnt/wingless signaling requires BCU9/legless-mediated recruitment of			
RT	pygopus to the nuclear beta-catenin-TCF complex.";			
RL	Cell 109:47-60(2002).			
CC	!- FUNCTION: Involved in signal transduction through the wnt pathway.			
CC	!- SUBUNIT: Binds to beta-catenin (CTNBL1), PYGO1 and PYGO2.			
CC	!- TISSUE/CELLULAR LOCATION: Nuclear (Probable).			
CC	!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,			
CC	testis, ovary and small intestine, and at lower levels in spleen,			
CC	colon and blood.			
CC	!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation			
CC	found in a patient with precursor B-cell acute lymphoblastic			
CC	leukemia (ALL). This translocation leaves the coding region			
CC	intact but may have pathogenic effects due to alterations in the			
CC	expression level of BCU9. Several cases of translocations within			
CC	the 3' untranslated region of BCU9 have been found in B-cell			
CC	thymomas.			
CC	!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.			
CC	!- CAUTION: Ref.1 sequence differs from that shown due to a			
CC	frameshift in position 1391.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			

Q99un9 staphylococ  
P34934 sus. scrofa  
P47040 saccharomyc  
O08032 saccharomyc  
P43439 enterococcu  
O02949 oviss aries  
Q8C4D2 mechanopyru  
P33059 saccharomyc  
P44934 haemophilus  
Q8R424 thermoaer  
P58790 agrobacteri  
Q91984 bacillus st

34 48 27.0 328 1 PLSX STRAM  
35 48 27.0 379 1 HSYX BIG  
36 48 27.0 408 1 EYNI YEAST  
37 48 27.0 650 1 CC45 YEAST  
38 48 27.0 664 1 NTF1 ENTHR  
39 48 27.0 664 1 ED13 SHEEP  
40 47.5 26.7 309 1 PYRS METWA  
41 47.5 26.7 558 1 MUT2 YEAST  
42 47.5 26.7 861 1 MUT2 HAEIN  
43 47 26.4 184 1 RRF THETN  
44 47 26.4 247 1 HIS4 AGRTS  
45 47 26.4 296 1 HSLQ BACST

EMBL; Y13620; CAA73942.1; ALT\_FRAME.  
 DR Genew; HGNC:1008; BCL9.  
 DR MIW; 602597; -  
 DR GO; GO:0007048; P:oncogenesis; TAS.  
 DR Nuclear protein; Chromosomal translocation; Proto-oncogene;  
 KW Wnt signaling pathway.  
 FT DOMAIN 331 1378 PRO-RICH.  
 FT DOMAIN 347 377 CYNBI-BINDING.  
 FT DOMAIN 331 335 POLY-PRO 1.  
 FT DOMAIN 514 517 POLY-PRO 2.  
 FT DOMAIN 900 903 POLY-ALA.  
 FT DOMAIN 970 973 POLY-PRO 3.  
 FT DOMAIN 973 973 POLY-PRO 3.  
 SQ SEQUENCE 1426 AA; 149314 MW; A240A48771687F1B CRC64;  
 Query Match 100.0%; Score 178; DB 1; Length 1426;  
 Best Local Similarity 100.0%; Pred. No. 3; le-15;  
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGLSQQLHRSLSQTLRDIQRMFLPFDEKFTCA 35  
 DB 349 DGLSQQLHRSLSQTLRDIQRMFLPFDEKFTCA 349  
 RESULT 2  
 BCL9-DROME STANDARD; PRT; 1469 AA.  
 ID BCL9-DROME  
 AC Q961D9; Q9V4D2; -  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE BCL-9 homolog (Legless protein).  
 GN BCL9 OR LGS OR CG2041.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyridae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN BCL9-DROME  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=berkeley; PubMed=10731132;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams W.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Calle R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Landell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.A.C., Blaise R.G., Change M., Felfel B.D.,  
 RA Wan K.H., Doyle E.G., Heit G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.P., Agbayani A., An H.-J., Andrews-Franko C., Baldwin D.,  
 RA Ballew R.M., Basu A., Bavendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Besson K.I., Benos P., Bernier B.P., Bhandari P., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck H., Bockstein P., Bottler P.,  
 RA Burtis K.C., Buzan D.N., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo E., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Garg R.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaiswal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.B., Kodra C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McInosh T.C., McLeod W.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy P., Murphy L., Musny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.N., Weissbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 Zhao X.H., Zhong F.N., Zhong R.N., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 Glibos R.A., Myers E.W., Rubin G.M., Rubin J.C.,  
 "The genome sequence of Drosophila melanogaster";  
 Science 287:2185-2195(2000).  
 [2]  
 SEQUENCE FROM N.A.  
 RP STRAIN=berkeley; TISSUE=Embryo;  
 RC MEDLINE=22426066; PubMed=12537569;  
 RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champagne M.,  
 RA George R.A., Garin H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E., CNA resource";  
 "A Drosophila full-length cDNA resource";  
 Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 [3]  
 SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-514 AND  
 ILE-513.  
 RP MEDLINE=21952490; PubMed=11955446;  
 RX Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
 RA Marone M., Zuehlig S., Basler K., BCL9/Legless-mediated recruitment of  
 RA "Wnt/Wingless signaling requires BCL9/Legless-mediated recruitment of  
 RA Pygopus to the nuclear beta-catenin-TCF complex";  
 Cell 109:47-60(2002).  
 RL -1- FUNCTION: Involved in signal transduction through the wnt pathway.  
 CC -1- SUBUNIT: Binds to ARM and PYGO.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 DR EMBL; AE003844; AF59345.1; ALT\_SEQ.  
 DR EMBL; AY01851; AK93075.1; -  
 DR EMBL; AF457205; AAL91468.1; -  
 DR FlyBase; FBgn0039907; Lgs.  
 DR GO; GO:0005634; C:nucleus; IE.  
 DR GO; GO:000528; P:transcription regulator activity; IPI.  
 DR GO; GO:003077; P:positive regulation of wnt receptor signaling; IPI.  
 DR GO; GO:0007367; P:segment polarity determination; IMP.  
 KW Nuclear protein; Developmental protein; Segmentation polarity protein;  
 KW Wnt signaling pathway.  
 FT DOMAIN 511 555 ARM-BINDING.  
 FT DOMAIN 1134 1173 ASN-RICH.  
 FT DOMAIN 1340 1449 GLN-RICH.  
 FT DOMAIN 1162 1169 POLY-ASN.  
 FT MUTAGEN 514 514 G->E: IN ALLELE LGS-21L.  
 FT MUTAGEN 534 534 L->F: IN ALLELE LGS-17E; SEGMENT POLARITY  
 FT PHENOTYPE.  
 FT MUTAGEN 537 537 1->K: IN ALLELE LGS-17P.  
 SQ SEQUENCE 1469 AA; 153759 MW; 5672E01B7200ED08 CRC64;  
 Query Match 37.1%; Score 66; DB 1; Length 1469;  
 Best Local Similarity 31.4%; Pred. No. 1.2;  
 Matches 11; Conservative 10; Mismatches 14; Indels 0; Gaps 0;  
 QY 1 DGLSQQLHRSLSQTLRDIQRMFLPFDEKFTCA 35  
 DB 520 ENLTPOQRHESQALAKRMNQLPFENESVGA 554  
 RESULT 3  
 TRF4 YEAST  
 ID TRF4 YEAST  
 AC PS3632; STANDARD; PRT; 584 AA.

```

DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Topoisomerase 1-related protein TRP4.
GN TRP4 OR YOL115W OR O0716 OR HRC584.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96109595; PubMed=8647385;
RA Sadoff B.U., Heath-Pagiluso S., Castano I.B., Zhu Y., Kieff P.S.,
RT "Isolation of mutants of Saccharomyces cerevisiae requiring DNA
RT topoisomerase I."
RL Genetics 141:465-479 (1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076631; PubMed=7502582;
RA Vandenbol M., Durand P., Portetelle D., Hilger F.;
RT "Sequence analysis of a 44 kb DNA fragment of yeast chromosome XV
RT including the Tyl-H3 retrotransposon, the sufl1(+) frameshift
RT suppressor gene for tRNA-Gly, the yeast transfer RNA-Thr-1a and a
RT delta element."
RL Yeast 11:1069-1075 (1995).
CC -1- FUNCTION: ESSENTIAL PROTEIN REQUIRED FOR PROPER NUCLEAR DIVISION
CC -1- IN MITOSIS. MAY MEDIATE MITOTIC CHROMOSOME CONDENSATION.
CC -1- SIMILARITY: BELONGS TO THE CID1/TRF4/TRF5 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
CC EMBL: U11355; AAC49091.1;
CC EMBL: Z48149; CA86145.1;
CC EMBL: Z74857; CA93134.1;
CC PIR: S51882; S51882.
CC SGD: S0003475; TRP4.
CC GO: GO:0005634; C:nucleus; IDA.
CC GO: GO:0003887; F:DNA-directed DNA polymerase activity; IDA.
CC GO: GO:0006285; P:DNA topological change; IGI.
CC GO: GO:0007076; P:mitotic chromosome condensation; IMP.
CC InterPro: IPR002934; NTP_transf.
CC InterPro: IPR001201; PAP_25A_core.
CC InterPro: IPR002058; PAP_assoc.
CC Pfam: PF01909; NTP_transf_2; 1.
CC Pfam: PF03828; PAP_assoc; 1.
CC MitoSis.
SQ SEQUENCE 584 AA; 66030 MW; 8A58B29E4BFC022 CRC64;

Query Match 31.5%; Score 56; DB 1; Length 584;
Best Local Similarity 36.0%; Pred. No. 8.7;
Matches 9; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 4 SQEQLHRRSRSLQTLRIQRMFPD 28
DB 193 SREBIEIRNQITSTIREAVKQLWPD 217

RESULT 4
HEMI_SYNEL
ID HEMI_SYNEL STANDARD; PRT; 426 AA.
AC Q8DIE3;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
GN HENA OR TLL1738.

OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BP-1;
RX MEDLINE=22255144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Kato H., Sasamoto S.,
RA Watanabe A., Iriuchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Takeuchi C., Yamada M., Tabata S.;
RA Shimo S., Sugimoto M., Matsuno A., Matsuda M., Yamada M.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130 (2002).
CC -1- CATALYTIC ACTIVITY: Glutamyl-tRNA (Glu) + NADPH = glutamate-1-
CC semialdehyde + NADP(+) + tRNA (Glu).
CC -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
CC Involved in chlorophyll biosynthesis.
CC -1- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenses@isb-sib.ch).
CC
CC EMBL: AP005375; BAC09290.1;
CC HAMAP: MF_00087; 1.
CC InterPro: IPR000343; GluTR.
CC InterPro: IPR000594; This domain.
CC Pfam: PF00745; GluTR_dimer; 1.
CC Pfam: PF05201; GluTR_N; 1.
CC Pfam: PF05200; GluTR_NAD_bind; 1.
CC TIGRfam: TIGR01035; "hemA_1".
CC PROSITE: PS00747; GluTR; FALSE NEG.
CC Porphyrin biosynthesis; Chlorophyll biosynthesis; Oxidoreductase;
CC NADP; Complete proteome.
CC ACT_SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
CC ACT_SITE 99 99 BASE (BY SIMILARITY).
CC SEQUENCE 426 AA; 47596 MW; D84CE5A1D2AA77E CRC64;

Query Match 30.9%; Score 55; DB 1; Length 426;
Best Local Similarity 55.6%; Pred. No. 8.2;
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 4 SQEQLHRRSRSLQTLRI 21
DB 401 SQDLESQRQAWQLQDL 418

RESULT 5
RHGB_HUMAN
ID RHGB_HUMAN STANDARD; PRT; 718 AA.
AC Q9NSG0; O75983; O95695; O96RW1; O96RW2; O9HA49; Q9HC46; Q9NVX8;
AC Q9NXL1; Q9UHZ0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Rho-GTPase-activating protein 8.
GN RHGAP8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han T.H.;
RT "Novel human cDNA clone with function of inhibiting cancer cell
RT growth.";
```



```

CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ277605; CAC37689.1; -
CC EMBL; AB046448; BAB83911.1; -
CC EMBL; AJ400878; CAB92296.1; -
CC GSD; MGI:1928820; Asc13.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0005667; C:transcription factor complex; IPI.
CC GO; GO:0003677; P:DNA binding activity; IDA.
CC GO; GO:0005515; P:protein binding activity; IPI.
CC GO; GO:0030528; P:transcription regulator activity; IDA.
CC GO; GO:0006357; P:regulation of transcription from Pol II promoter; IDA.
CC InterPro; IPR001092; HLH_basic.
CC Pfam; PF00010; HLH_1.
CC SMART; SM00353; HLH_1.
CC PROSITE; PS00038; HLH_1; 1.
CC PROSITE; PS00888; HLH_2; 1.
CC Transcription regulation; Repressor; DNA-binding; Nuclear protein.
CC FT
CC DNA_BIND 95 105 BASIC DOMAIN.
CC DOMAIN 106 145 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
CC SEQ_ID 174 AA; 20245 MW; D89E56C8A9D3440B CRC64;
CC -----
Query Match 30.6%; Score 54.5; DB 1; Length 174;
Best Local Similarity 36.1%; Pred. No. 3.5;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;
-----
QY 3 LSQEQLEHRSLSQTLRD-----IQRLPDPKE 31
Db 119 LPEDYLEKRSKVTLRRAAIKYISYLSLLYPDESE 154
-----
RESULT 7
PHYB_ORYSA
ID PHYB_ORYSA STANDARD; PRT; 1171 AA.
AC P25764;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DR PHYTOCHROME_B.
GN PHYB OR PHYB1
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhacidae; Oryzaceae; Oryza.
CC NCBI_TaxID=4530;
CC [1]
CC SEQUENCE FROM N.A.
RP STRAIN=cv. Indica-IR36; TISSUE=Seedling shoot;
RX MEDLINE=91172131; PubMed=2005872;
RA Devesh K., Tepperman J., Christensen A.H., Quail P.H.;
RT "phyb is evolutionarily conserved and constitutively expressed in
RT rice seedling shoots.";
RL Mol. Gen. Genet. 225:305-313(1991).
CC -1- FUNCTION: REGULATORY PHOTORECEPTOR WHICH EXISTS IN TWO FORMS THAT
CC ARE REVERSIBLY INTERCONVERTIBLE BY LIGHT. THE PR FORM THAT ABSORBS
CC MAXIMALLY IN THE RED REGION OF THE SPECTRUM AND THE PFR FORM THAT
CC ABSORBS MAXIMALLY IN THE FAR-RED REGION. PHOTOCOVERSION OF PFR IN
CC PFR INDUCES AN ARRAY OF MORPHOGENIC RESPONSES. WHEREAS
CC RECONVERSION OF PFR TO PR CANCELS THE INDUCTION OF THOSE
CC RESPONSES. PFR CONTROLS THE EXPRESSION OF A NUMBER OF NUCLEAR
CC GENES INCLUDING THOSE ENCODING THE SMALL SUBUNIT OF RIBULOSE-
CC BISPHOSPHATE CARBOXYLASE, CHLOROPHYLL A/B BINDING PROTEIN,
CC PROTOCHLOROPHYLLIDE REDUCTASE, RNNA, ETC. IT ALSO CONTROLS
CC THE

```





[illegible]

```

GN ASCL3 OR SGN1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S, Ohno K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.
RT "Sgn1, a basic helix-loop-helix transcription factor delineates the
RT salivary gland duct cell lineage in mice."
RL Dev. Biol. 240:517-530(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21418998; PubMed=11528127;
RA Amold C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
RT architecture of a conserved syntenic region on human chromosome
RT 1p15.3 (including gene STS) and mouse chromosome 7."
RL Cytogenet. Cell Genet. 93:284-290(2001).
CC -!- FUNCTION: Transcriptional repressor. Inhibits myogenesis (by
CC similarity).
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC bHLH protein (by similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB046450; BAB83913.1;
DR EMBL; AJ408877; CAB82288.1; ALT_INIT.
DR Genew; HGNC:740; ASCL3.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA BIND 95 105
FT DOMAIN 106 145
FT CONFLICT 54 54 R -> L (IN REF. 1).
SQ SEQUENCE 180 AA; 20796 MW; C9B9C885DE94F81 CRC64;
Query Match 29.5%; Score 52.5; DB 1; Length 180;
Best Local Similarity 36.1%; Pred. No. 6.6;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;
QY 3 LSQQLHRRSRLQTRD-----IQRLFPDEKE 31
DB 119 LPEEYLEKLSKVETLRAAIKVINYLSLLYDPAE 154
RESULT 12
RP RRF RUSNN
ID - RRF RUSNN STANDARD; PRT; 190 AA.
AC Q8R5Z5; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
GN FRR OR FN1623.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S, Ohno K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.
RT "Sgn1, a basic helix-loop-helix transcription factor delineates the
RT salivary gland duct cell lineage in mice."
RL Dev. Biol. 240:517-530(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21418998; PubMed=11528127;
RA Amold C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
RT architecture of a conserved syntenic region on human chromosome
RT 1p15.3 (including gene STS) and mouse chromosome 7."
RL Cytogenet. Cell Genet. 93:284-290(2001).
CC -!- FUNCTION: Transcriptional repressor. Inhibits myogenesis (by
CC similarity).
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC bHLH protein (by similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB046450; BAB83913.1;
DR EMBL; AJ408877; CAB82288.1; ALT_INIT.
DR Genew; HGNC:740; ASCL3.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA BIND 95 105
FT DOMAIN 106 145
FT CONFLICT 54 54 R -> L (IN REF. 1).
SQ SEQUENCE 180 AA; 20796 MW; C9B9C885DE94F81 CRC64;
Query Match 29.5%; Score 52.5; DB 1; Length 180;
Best Local Similarity 36.1%; Pred. No. 6.6;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;
QY 3 LSQQLHRRSRLQTRD-----IQRLFPDEKE 31
DB 119 LPEEYLEKLSKVETLRAAIKVINYLSLLYDPAE 154
RESULT 12
RP RRF RUSNN
ID - RRF RUSNN STANDARD; PRT; 190 AA.
AC Q8R5Z5; 2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ribosome recycling factor (Ribosome releasing factor) (RRF).
GN FRR OR FN1623.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
OC Fusobacterium.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S, Ohno K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.
RT "Sgn1, a basic helix-loop-helix transcription factor delineates the
RT salivary gland duct cell lineage in mice."
RL Dev. Biol. 240:517-530(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21418998; PubMed=11528127;
RA Amold C., Bahr A., Mujica A., Sampson N., Bikar S.E., Winterpacht A.,
RA Zabel B., Hankeln T., Schmidt E.R.;
RT "Comparative genomic sequencing reveals a strikingly similar
RT architecture of a conserved syntenic region on human chromosome
RT 1p15.3 (including gene STS) and mouse chromosome 7."
RL Cytogenet. Cell Genet. 93:284-290(2001).
CC -!- FUNCTION: Transcriptional repressor. Inhibits myogenesis (by
CC similarity).
CC -!- SUBUNIT: Efficient DNA binding requires dimerization with another
CC bHLH protein (by similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS. ASC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB046450; BAB83913.1;
DR EMBL; AJ408877; CAB82288.1; ALT_INIT.
DR Genew; HGNC:740; ASCL3.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
FT DNA BIND 95 105
FT DOMAIN 106 145
FT CONFLICT 54 54 R -> L (IN REF. 1).
SQ SEQUENCE 180 AA; 20796 MW; C9B9C885DE94F81 CRC64;
Query Match 29.5%; Score 52.5; DB 1; Length 190;
Best Local Similarity 36.1%; Pred. No. 7;
Matches 13; Conservative 8; Mismatches 10; Indels 5; Gaps 1;
QY 3 LSQQLHRRSRLQTRD-----IQRLFPDEKEFT 33
DB 153 ISEDLKKEINVTQLTKVVKEDILLAKKEKIT 188
RESULT 13
ID - Y755 AQUAE STANDARD; PRT; 242 AA.
AC O66957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_755.
GN AQ_755.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```
RT cyanobacterium Anabaena sp. strain PCC 7120." ;
RL DNA Res. 8:205-213(2001).
CC -!- FUNCTION: THE SIGMA FACTOR IS AN INITIATION FACTOR THAT PROMOTES
CC ATTACHMENT OF THE RNA POLYMERASE TO SPECIFIC INITIATION SITES AND
CC THEN IS RELEASED.
CC -!- SIMILARITY: Belongs to the sigma-70 factor family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M95760; AAA22046.1; -.
DR EMBL; AP003602; BAB77258.1; -.
DR PIR; AD2541; AD2541.
DR HSSP; P00579; ISIG.
DR InterPro; IPR000943; Sigma_70.
DR Pfam; PF00140; sigma70_r1_2; 1.
DR Pfam; PF04542; sigma70_r2; 1.
DR Pfam; PF04539; sigma70_r3; 1.
DR Pfam; PF04545; sigma70_r4; 1.
DR PRINTS; PR00046; SIGMA70FCT.
DR PROSITE; PS00715; SIGMA70_1; 1.
DR PROSITE; PS00716; SIGMA70_2; 1.
DR Transcription regulation; Sigma factor; DNA-directed RNA polymerase;
KW DNA-binding; Plasmid; Complete proteome.
FT DOMAIN 125 138 POLYMERASE CORE BINDING (POTENTIAL).
FT DNA_BIND 294 313 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 332 AA; 38431 MW; 81323E670C6DF46A CRC64;

Query Match 29,2%; Score 52; DB 1; Length 332;
Best Local Similarity 35,5%; Pred. No. 15;
Matches 11; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

QY 1 DGLSQQLERHERSLQTLRDIQRLFPDEKE 31
DB 248 DMSPEFAERRELLYQDIHNLAKLTQQKE 278
```

Search completed: November 13, 2003, 09:25:59  
Job time : 5.1323 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 11.4397 Seconds  
(Without alignments)  
789.518 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_349\_383

Perfect score: 178

Sequence: 1 DGLSQEQLHRRSLQTLRIQRMFLPDKKEFTGA 35

Scoring table: BLOSUM62

Gap: 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

- 1: sp.archaea:\*
- 2: sp.bacteria:\*
- 3: sp.fungi:\*
- 4: sp.human:\*
- 5: sp.invertebrate:\*
- 6: sp.mammal:\*
- 7: sp.muc:\*
- 8: sp.organelle:\*
- 9: sp.plant:\*
- 10: sp.plant:\*
- 11: sp.rodent:\*
- 12: sp.virus:\*
- 13: sp.vertibrate:\*
- 14: sp.unclassified:\*
- 15: sp.rvitus:\*
- 16: sp.bacteriap:\*
- 17: sp.archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	32.0	1034	16	Q828Y6
2	57	32.0	1046	16	Q8ZRE3
3	56	31.5	859	4	Q9NT51
4	56	31.5	1132	4	Q9V2H2
5	56	31.5	1208	5	Q9VXU1
6	56	31.5	1398	5	Q9VXU2
7	55	30.9	295	16	Q8UCAO
8	55	30.9	426	16	Q8DI53
9	55	30.9	433	4	Q8IZM6
10	55	30.9	818	17	Q8TG21
11	53.5	30.1	155	11	Q8VD56
12	53.5	30.1	4957	4	O14687
13	53.5	30.1	5262	4	O14686
14	53	29.8	227	16	O87532
15	53	29.8	411	16	Q8PRD9
16	52.5	29.5	673	5	O01302

17	52	29.2	399	10	Q8W0M3
18	52	29.2	572	16	Q9HT07
19	52	29.2	616	12	P87708
20	52	29.2	946	4	Q9SQE4
21	52	29.2	1009	3	Q06407
22	52	29.2	1766	12	Q91599
23	52	29.2	2607	11	Q8RT18
24	51.5	28.9	1146	10	Q9F802
25	51.5	28.9	1151	10	Q9F803
26	51	28.7	102	16	Q8ZDE6
27	51	28.7	108	16	Q8D0S6
28	51	28.7	239	16	Q8FFZ8
29	51	28.7	258	11	Q8SVF0
30	51	28.7	329	10	Q8UG66
31	51	28.7	329	10	Q9CG75
32	51	28.7	337	5	O62022
33	51	28.7	341	4	O81YY3
34	51	28.7	449	5	Q810M4
35	51	28.7	848	16	Q8FMB8
36	51	28.7	1047	16	Q8XE36
37	51	28.7	1464	5	Q23995
38	51	28.7	1464	5	Q9VC47
39	51	28.7	1464	5	Q24132
40	51	28.7	1755	15	O56220
41	51	28.7	1755	15	O91ZT3
42	50.5	28.4	96	5	Q9VN78
43	50.5	28.4	288	13	Q8UVZ1
44	50.5	28.4	657	10	Q9FN06
45	50.5	28.4	675	4	Q9HD27

## ALIGNMENTS

### RESULT 1

Q828Y6	PRELIMINARY;	PRT; 1034 AA.
ID	Q828Y6	
AC	Q828Y6;	
DT	01-MAR-2002 (TRENBLER, 20, Created)	
DT	01-MAR-2002 (TRENBLER, 20, Last sequence update)	
DT	01-MAR-2003 (TRENBLER, 23, Last annotation update)	
DE	Exonuclease SbCC.	
GN	STY0429.	
OS	Salmonella typhi.	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
OC	Enterobacteriaceae; Salmonella.	
OX	NCBI_TaxID=601;	
RN	(1)	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CT18;	
RX	MEDLINE=21534947; PubMed=11677608;	
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,	
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,	
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,	
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,	
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,	
RA	Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,	
RA	Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,	
RA	Whitehead S., Barrell B.G.;	
RT	"Complete genome sequence of a multiple drug resistant Salmonella	
RT	enterica serovar Typhi CT18.";	
RL	Nature 413:848-852(2001).	
DR	EMBL: AL627266; CAP08850.1; ..	
DR	InterPro: IPR003439; ABC transporter.	
DR	InterPro: IPR004592; SbcC.	
DR	InterPro: IPR003395; SMC N.	
DR	Pfam: PF02463; SMC N; 1.	
KW	TIGRFAMS: TIGR00618; sbcc; 1.	
SQ	Complete proteome; sbcc; 1.	
SQ	SEQUENCE 1034 AA; 116759 MW; 2513B7573626960A CRC64;	

Query Match 32.0%; Score 57; DB 16; Length 1034;  
Best Local Similarity 42.9%; Pred. No. 68;

Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQQLHEHRSLSQTLDIQRLMFPDEK 30

Db 213 LADQLQLEASLQALTDKRLADQ 240

# RESULT 2

Q82RE3 AC Q82RE3 PRELIMINARY; PRT; 1046 AA.  
 DT 01-NAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-NAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-NAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE ATP-dependent deRNA exonuclease.  
 GN SBCC OR STM0395.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=924112 / RCSC 700720;  
 RA MEDLINE=2154948; PubMed=11677609;  
 RA McClelland M., Sanderson K.A., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney S., Porwollik S., Ali J., Dantas M., Da P., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking L., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RA "Complete genome sequence of *Salmonella enterica* serovar Typhimurium  
 LT2";  
 RL Nature 413:852-856 (2001).  
 DR EMBL; AB008713; AL019349.1; --  
 DR InterPro; IPR003439; ABC transporter.  
 DR InterPro; IPR004592; SMC.  
 DR InterPro; IPR003395; SMC N.  
 DR Pfam; PF02463; SMC\_N; 1.  
 DR TIGRFAMs; TIGR00618; sbcc; 1.  
 KW Exonuclease; Complete proteome.  
 SQ SEQUENCE 1046 AA; 117824 MW; BA565CA3BDAD0C82 CRC64;

Query Match 32.0%; Score 57; DB 16; Length 1046;  
 Best Local Similarity 42.9%; Pred. No. 69;  
 Matches 12; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 3 LSQQLHEHRSLSQTLDIQRLMFPDEK 30

Db 213 LADQLQLEASLQALTDKRLADQ 240

# RESULT 3

Q82RE3 AC Q82RE3 PRELIMINARY; PRT; 859 AA.  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN DKEP2434P1818.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=testis;  
 RA O'Brien B., Obermaier B., Mewes H.W., Gassenhuber J., Wiemann S.;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL137528; CAB70792.1;  
 DR InterPro; IPR002013; Syja N.  
 DR Pfam; PF02383; Syja N; 1.  
 DR PROSITE; PS50275; SAC; 1.  
 KW Hypothetical protein.

FT NON\_TER 1 1  
 SQ SEQUENCE 859 AA; 96781 MW; BCBC47C8B726D76 CRC64;

Query Match 31.5%; Score 56; DB 4; Length 859;  
 Best Local Similarity 33.3%; Pred. No. 76;  
 Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLEHRSLSQTLDIQRLMFPDEKFTG 34

Db 317 ENQSHQELISQLSYMKLLLPDDEKPHG 346

# RESULT 4

Q9V2H2 AC Q9V2H2 PRELIMINARY; PRT; 1132 AA.  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
 DE Hypothetical protein KIAA0966.  
 GN KIAA0966.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue=Brain;  
 RX MEDLINE=99246063; PubMed=10231032;  
 RA Nagai T., Ishikawa K., Suyama M., Kikuno R., Hirosewa M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RA "Prediction of the coding sequences of unidentified human genes. XIII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro";  
 RL DNA Res. 6:63-70 (1999).  
 DR EMBL; AB03183; BAA76810.1; --  
 DR InterPro; IPR002013; Syja N.  
 DR Pfam; PF02383; Syja N; 1.  
 DR PROSITE; PS50275; SAC; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 1132 AA; 128407 MW; 853719FC0AD455CD CRC64;

Query Match 31.5%; Score 56; DB 4; Length 1132;  
 Best Local Similarity 33.3%; Pred. No. 1e+02;  
 Matches 10; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 5 QEQLEHRSLSQTLDIQRLMFPDEKFTG 34

Db 590 ENQSHQELISQLSYMKLLLPDDEKPHG 619

# RESULT 5

Q9VXU1 AC Q9VXU1 PRELIMINARY; PRT; 1208 AA.  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE CG32587 protein.  
 GN CG32587 OR CG6350 OR CG7821.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=serkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

Query Match 31.5%; Score 56; DB 5; Length 1208;  
Best Local Similarity 45.5%; Pred. No. 1.e+02;  
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps

3 ISOFOLEHERRESIQTILRIQFORM 24  
|||..|||.....  
811 LQQQAQSQQCAQSTLRDLREL 832

RESULT 6  
SEQUENCE  
Q9VXU2 PRELIMINARY; PRT; 1398 AA.  
Q9VXU2; Q9G0D0;  
01-MAY-2000 (TRENBUrel. 13, Created)  
01-OCT-2002 (TRENBUrel. 22, Last sequence update)  
01-MAR-2003 (TRENBUrel. 23, Last annotation update)  
CG32587 protein (S007366p).  
CG32587 OR CG6350 OR CG7821.  
Drosophila melanogaster (Fruit fly).  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
NCBI TaxID=7237;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=Berkeley;  
MEDLINE=20196006; PubMed=10731132;  
Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Gallie R.P.,  
George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
Sutton G.C., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
Brandon R.G., Rogers J.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
Wan K.H., Doyle C., Baxter E.C., Helt G., Nelson C.R., Miklos G.L.G.,  
Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
Ballou R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,  
Beeson X.Y., Benos P.V., Bereman B.P., Bhattacharya D., Bolshakov S.,  
Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P.,  
Burry K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
Cherry J.M., Cawley S., Dahlke C., Davenport L.S., Davies P.,  
de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
Flores C., Gabrielian A.B., Gang N.S., Gelbart W.M., Glasser K.,  
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Haeris W.,  
Haas R., Hasty D., Heiman T.J., Hernandez J.P., Houck C.,  
Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,  
Jailai M., Kaulous F., Karpen G.H., Ke Z., Li Z., Liang Y., Lin X.,  
Kimmel B.E., Kodira C.D., Kraft C., Li J., Liu Z., Liang Y., Lin X.,  
Lasko P., Lei Y., Levitsky A.C., Li J., Liu Z., Liang Y., Lin X.,  
Lin X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,  
Merkulov G., Milshina N.V., Mobarry C., Mulvey D.M., Nelson D.L.,  
Nelson D.R., Nelson K.A., Nixon K., Murphy L., Murty D.M., Nelson D.L.,  
Palazzo M., Pittman G.S., Pan S., Pollard J.R., Puri V., Reese M.G.,  
Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
Shue B.C., Siden-Kiamos I., Simpson M., Stokely M.P., Smith T.,  
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
Swirskis R., Tecator C., Turner R., Venter E., Wang A.H., Wang X.,  
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
Ye J., Ye R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
"The genome sequence of Drosophila melanogaster.";  
Science 287:2185-2195(2000).

Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
Barnson J., An H., Baldwin D., Batton J., Beeson K.Y., Busam O.A.,  
Carlson J.W., Center A., Champs E., Champagne L.B., Dietz S.M.,  
Dodson K., Dorsett V., Doull L.E., Doyle C., Dresnek D., Farfan D.,



RA Ferriera S., Friese E., Galle R.F., Garg N.S., George R.A.,  
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
RA MacIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
RA Pacleb J., Paragov S., Park S., Patel S., Pfeiffer B.,  
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
RL "Sequencing of *Drosophila melanogaster* genome";  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RN SEQUENCE FROM N.A.  
RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
RP Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
RA Tupy J.L., Beigman C., Berniker B., Carlson J.W., Ceiniker S.E.,  
RA Klamp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,  
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,  
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
RT "Annotation of *Drosophila melanogaster* genome";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RN SEQUENCE FROM N.A.  
RP Adams M.D., Ceiniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RN SEQUENCE FROM N.A.  
RP FlyBase;  
RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RN SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Friese E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nunoo J., Pacleb J., Paragov S., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF003499; AF48466.2; -;  
DR EMBL; AF052121; AXK93545.1; -;  
DR FlyBase; FBN0052587; CG32587.  
SQ SEQUENCE 1398 AA; 158484 MW; 8BA1C3FE3E9F55D CRC64;

Query Match 31.5%; Score 56; DB 5; Length 1398;  
Best Local Similarity 45.5%; Pred. No. 1.2e+02;  
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

OY 3 LSGQLHRSLSLOTLDYORM 24  
DB 1001 LQQQAESQEQASTLDRL 1022

RESULT 7  
ID QBUCAO PRELIMINARY; PRT; 295 AA.  
AC QBUCAO;  
DT 01-JUN-2002 (TEMBLrel. 21, Created)  
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)  
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
DE Transcriptional regulator, RpiR family.  
GN ATU2598 OR AGR\_C 4707.  
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium.  
OX NCBI\_TaxID=176299;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP Wood D.W., Serubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
RA Wood D.W., Serubal J.C., Kaul R., Monks D.E., Almeida N.F. Jr., Woo L.,  
RA Chou Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sk.,  
RA Chapman J., Glendinning J., Deatherage G., Gillet W., Grant C.,  
RA Kutayin T., Levy R., La W.-J., McClelland E., Palmeri A.,  
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,

RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,  
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
RA Nester E.W.;  
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens C58";  
RL Science 294:2317-2323 (2001).  
RN [2]  
RN SEQUENCE FROM N.A.  
RP Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
RX MEDLINE=21608551; PubMed=11743194;  
RA Quorllo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
RA Houmle K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,  
RA Flanagan C., Crowell C., Gursen J., Lomo C., Gear C., Strub G.,  
RA Cleo C., Slater C.;  
RT "Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58";  
RL Science 294:2323-2328 (2001).  
DR EMBL; AF009206; AAL43579.1; ALT\_INIT.  
DR EMBL; AF008173; AAK88321.1; -;  
DR InterPro; IPR000843; HTH\_Laci.  
DR InterPro; IPR000281; HTH\_RpiR.  
DR InterPro; IPR001347; SIS.  
DR Pfam; PF01418; HTH\_6.1.  
DR Pfam; PF01380; SIS\_1.  
DR PROSITE; PS00356; HTH\_LACI\_FAMILY; 1.  
KW Complete proteome.  
SQ SEQUENCE 235 AA; 32259 MW; 385092683CE72456 CRC64;

Query Match 30.9%; Score 55; DB 16; Length 295;  
Best Local Similarity 40.0%; Pred. No. 35;  
Matches 12; Conservative 7; Mismatches 7; Indels 4; Gaps 1;

OY 9 EHRERSLOTLDYORMLF-----PDEKFTG 34  
DB 265 QQRQSWTLRHQKQVHRDPDDKQLG 294

RESULT 8  
ID QBD153 PRELIMINARY; PRT; 426 AA.  
AC QBD153;  
DT 01-MAR-2003 (TEMBLrel. 23, Created)  
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)  
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)  
DE Transfer RNA-Gln reductase.  
GN HENA OR TLL1738.  
OS Synecococcus elongatus (Thermosynechococcus elongatus).  
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.  
OX NCBI\_TaxID=32046;  
RN [1]  
RN SEQUENCE FROM N.A.  
RP STEIN=BP-1;  
RC MEDLINE=2225144; PubMed=12240834;  
RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,  
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,  
RA Kiyokawa C., Kohata M., Matsumoto M., Matsuno A., Nakazaki N.,  
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;  
RT "Complete genome structure of the thermophilic cyanobacterium Thermosynechococcus elongatus BP-1";  
RL DNA Res. 9:123-130 (2002).  
DR EMBL; AF005375; BAC09290.1; -;  
KW Complete proteome.  
SQ SEQUENCE 426 AA; 47596 MW; D84CE5A1D2AA777E CRC64;

Query Match 30.9%; Score 55; DB 16; Length 426;  
Best Local Similarity 55.6%; Pred. No. 50;  
Matches 10; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 4 SSOQLHRSLSLOTLDY 21  
DB 401 SQRDSQRANMQLQDL 418

```

RESULT 9
Q812M6 PRELIMINARY; PRT; 433 AA.
AC Q812M6;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE BCH domain-containing Cdc42GAP-like protein.
GN BFGAP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Shang X., Zhou Y.T., Low B.C.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF544240; AAN04769.1; -.
SQ SEQUENCE 433 AA; 49691 MW; 0AE4942A404AE1D3 CRC64;

Query Match 30.9%; Score 55; DB 4; Length 433;
Best Local Similarity 52.2%; Pred. No. 51;
Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GLSQEQLHRSRLQTLRIQRM 24
Dd 224 GLRTEGLPRASVQTVREIQRL 246

RESULT 10
Q8T3Z1 PRELIMINARY; PRT; 818 AA.
AC Q8T3Z1;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TEMBLrel. 22, Last annotation update)
DE Archaea-specific superfamily II helicase.
GN MK0835.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mazhevaeva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natile D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Setter K.O.,
RA Malysk A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19
RT and monophyly of archaeal methanogens."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).
DR EMBL; AE010374; AAM02048.1; -.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00382; AAA; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 818 AA; 91715 MW; C2136200A710817E CRC64;

Query Match 30.9%; Score 55; DB 17; Length 818;
Best Local Similarity 54.2%; Pred. No. 98;
Matches 13; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 2 GLSQEQLHRSRLQTLRIQRM 25
Dd 224 GLRTEGLPRASVQTVREIQRL 246

```

---

```

DB 134 GFSQTLLEKLERLHELDRV 157

RESULT 11
Q8VDS6 PRELIMINARY; PRT; 155 AA.
ID Q8VDS6;
AC Q8VDS6;
DT 01-MAR-2002 (TEMBLrel. 20, Created)
DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Putative BHLH transcription factor (fragment).
GN SGN-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-Sprague-Dawley;
RX MEDLINE=21643927; PubMed=11784080;
RA Yoshida S., Ohbo K., Takakura A., Takebayashi H., Okada T., Abe K.,
RA Nabeshima Y.;
RT "Sgn1, a Basic Helix-Loop-Helix Transcription Factor Delimits the
RT Salivary Gland Duct Cell Lineage in Mice."
RL Dev. Biol. 240:517-530 (2001).
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC TRANSCRIPTION FACTORS.
DR EMBL; AB046449; BAB3912.1; -.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00048; HLH_1; 1.
DR PROSITE; PS00888; HLH_2; 1.
FT NON TER 1
SQ SEQUENCE 155 AA; 18146 MW; 6E8ABD1DA16B88D6 CRC64;

Query Match 30.1%; Score 53.5; DB 11; Length 155;
Best Local Similarity 36.1%; Pred. No. 28;
Matches 13; Conservative 7; Mismatches 9; Indels 7; Gaps 1;

Qy 3 LSQQLHRSRLQTLRD-----IQMLFDPDEK 31
Dd 100 LPEDYLEKLSKVETLRAAKIYSLQSLYPDETE 135

RESULT 12
ID Q14687 PRELIMINARY; PRT; 4957 AA.
AC Q14687;
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE ALR.
GN ALR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=97388474; PubMed=9247308;
RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
RA Canaan E.;
RT "Structure and expression pattern of human ALR, a novel gene with
RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560 (1997).
CC -1- SIMILARITY: CONTAINS 1 SET DOMAIN.
DR EMBL; AF010404; AAC51735.1; C.
DR InterPro; IPR003889; FTRICH_C.
DR InterPro; IPR003888; FTRICH_N.
DR InterPro; IPR000910; HMG_12_Box.

```



RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,  
 RA Pereira H.A., Rossi A., Seta J.A.D., Silva C., de Souza R.F.,  
 RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,  
 RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,  
 RA "Scubai J.C., Kitajima J.P.;  
 RT "Comparison of the genomes of two Xanthomonas pathogens with differing  
 host specificities";  
 RL Nature 417:459-463(2002).  
 DR EMBL; AF011625; AAM34916.1; -.  
 DR InterPro; IPR002886; Peptidase\_M37.  
 DR Pfam; PF01551; Peptidase\_M37; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 411 AA; 44337 MW; E193496126355426 CRC64;

Query Match 29.8%; Score 53; DB 16; Length 411;  
 Best Local Similarity 36.4%; Pred. No. 89;  
 Matches 12; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 2 GLSQQLHRSLSIQLRLDIQRMIFPDKKFTG 34  
 Db 17 GASQSQRETERKIQQLADELTKTISADRELEG 49

Search completed: November 13, 2003, 09:27:39  
 Job time : 14.4397 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 11.4397 Seconds  
(without alignments)  
388.502 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136  
Sequence: 1 VYFSTWANKAEAVLKGQVSTIVSFH 28

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DS seq length: 0

Maximum DS seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	1426	23	Human legless homo
2	136	100.0	1435	22	Human BCL9 homolog
3	129	94.9	140	22	Human polypeptide
4	98	72.1	1429	22	Drosophila melanog
5	98	72.1	1464	23	D. melanogaster lg
6	94	69.1	1494	23	Mouse beta-catenin
7	91	66.9	1114	23	Mouse ORFX protein
8	67	49.3	320	23	Mouse beta-catenin
9	65	47.8	113	22	Human polypeptide

10	52	38.2	360	21	AAAG33446	Zea mays protein f
11	52	38.2	448	21	AAAG33445	Zea mays protein f
12	52	38.2	509	21	AAAG33444	Zea mays protein f
13	51	37.5	1049	22	ABBA0387	Drosophila melanog
14	50	36.8	130	23	ABBA9793	Human polypeptide
15	50	36.8	365	22	ABG15088	Novel human diagno
16	50	36.8	621	21	ABG15740	Human aminoacyl-tr
17	50	36.8	631	21	ABBA3285	Human ORFX ORF3049
18	50	36.8	1063	23	ABBA08919	Human aminoacyl tr
19	49	36.0	187	23	ABBA6221	Bifidobacterium lo
20	49	36.0	984	24	ABG15883	Aspergillus fumiga
21	49	36.0	1088	24	ABG15889	Aspergillus fumiga
22	48	35.3	498	21	ABG17932	Arabidopsis thalia
23	48	35.3	498	21	ABG17932	Arabidopsis thalia
24	48	35.3	609	21	ABG17932	Arabidopsis thalia
25	48	35.3	609	21	ABG17932	Arabidopsis thalia
26	48	35.3	630	21	ABG17932	Arabidopsis thalia
27	48	35.3	630	21	ABG17932	Arabidopsis thalia
28	48	35.3	630	21	ABG17932	Arabidopsis thalia
29	47	34.6	886	21	AAAG40431	Arabidopsis thalia
30	47	34.6	905	21	AAAG39800	Arabidopsis thalia
31	47	34.6	1095	21	AAAG40430	Arabidopsis thalia
32	47	34.6	1114	21	AAAG39799	Arabidopsis thalia
33	47	34.6	1168	21	AAAG40429	Arabidopsis thalia
34	47	34.6	1187	21	AAAG39798	Arabidopsis thalia
35	45.5	33.5	341	22	ABBA0889	Grapevine leafroll
36	45	32.7	401	22	ABBA0889	Drosophila melanog
37	44.5	32.7	119	21	AAAG08467	Novel human diagno
38	44.5	32.7	137	21	AAAG08467	Arabidopsis thalia
39	44.5	32.7	152	21	AAAG08466	Arabidopsis thalia
40	44.5	32.7	160	21	AAAG08465	Arabidopsis thalia
41	44.5	32.7	170	21	AAAG51007	Arabidopsis thalia
42	44.5	32.7	170	21	AAAG51007	Arabidopsis thalia
43	44.5	32.7	171	21	AAAG51006	A. thaliana enviro
44	44.5	32.7	177	21	AAAG51006	Arabidopsis thalia
45	44.5	32.7	447	22	AAAG82903	S. epidermidis ope
			450	22	AAAG83066	S. epidermidis ope

## ALIGNMENTS

RESULT 1	AA871229	
ID	AA871229	standard; Protein; 1426 AA.
XX	AA871229	
XX	AA871229	
XX	AA871229	
DT	18-NOV-2002	(first entry)
XX	Human legless homologue lgs/bcl9 protein.	
DE	Human legless homologue lgs/bcl9 protein.	
XX	Legless; human; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;	
XX	tissue proliferation; tumour; cytosolic; cellular disorder; colon;	
KW	blood disorder; cancer; breast; head and neck cancer; brain; thyroid;	
XX	medulloblastoma; skin cancer; tissue regeneration; tissue repair.	
XX	Human sapiens.	
OS	Human sapiens.	
XX	US2002086986-A1.	
PN	US2002086986-A1.	
XX	04-JUL-2002.	
XX	04-JUL-2002.	
XX	27-JUL-2001; 2001US-0915543.	
PF	27-JUL-2001; 2001US-0915543.	
XX	28-JUL-2000; 2000US-221502P.	
PR	(BASL/) BASLER K.	
XX	(BRUN/) BRUNNER E.	
PA	(PROB/) FROESCH B.	
PA	(KRAM/) KRAMPS T.	
PA	(PETER/) PETER O.	
PI	Basler K, Brunner E, Froesch B, Kramps T, Peter O;	

XX WPI: 2002-63569/58.  
 XX N-PSDB; AAF98467.  
 XX Novel polypeptide useful in therapeutic method for treating disorders  
 XX of cell fate such as cell differentiation or cell proliferation  
 XX Example II; Fig 8B; 41pp; English.  
 XX This invention describes a novel polypeptide sharing one or more  
 XX homologous amino acid domains with the legless (lgs) protein, a  
 XX downstream component of the Wnt/Wingless (Wnt/Wg) signaling pathway  
 XX involved in the formation and maintenance of spatial arrangements  
 XX and proliferation of tissues during development, and in the formation  
 XX and growth of many human tumours. The products of the invention have  
 XX cytoskeletal activity and can be used to treat cellular disorders, blood  
 XX disorders and cancers caused by over-stimulation of the Wnt pathway.  
 XX where the cancerous condition is colon, breast, head and neck, brain,  
 XX thyroid, medulloblastoma or skin cancer. The product could also be used  
 XX to promote tissue regeneration and repair. This sequence represents the  
 XX human legless (lgs) protein homologue lgs/bcl9 described in the  
 XX disclosure of the invention.  
 XX Sequence 1426 AA;  
 XX Query Match 100.0%; Score 136; DB 23; Length 1426;  
 XX Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
 XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28  
 DB 177 VYVFSTEMANKAAEAVLKQGVETIVSFH 204  
 RESULT 2  
 ABB11808  
 ID ABB11808 standard; peptide; 1435 AA.  
 AC ABB11808;  
 DT 11-JAN-2002 (first entry)  
 XX Human BCL9 homologue, SEQ ID NO:2178.  
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;  
 XX haematopoietic regulation; tissue growth; immunomodulator; activin;  
 XX inhibin; chemotaxis; chemokinesis; cancer; tumour; haematopoietic disorder;  
 XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 XX myeloid cell disorder; lymphoid cell disorder; actin; arthritis;  
 XX chronic inflammatory condition; proliferative retinopathy;  
 XX atherosclerosis; coronary heart disease; arterial ischaemia;  
 XX bone disorder; osteoporosis; vascular growth disorder;  
 XX tissue regeneration; wound healing; infection; immune disorder;  
 XX cell culture; drug screening; gene therapy; anti-inflammation;  
 XX antischismatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 XX cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 XX antifungal; vulnery; antiulcer.  
 XX Homo sapiens.  
 XX WO200157188-A2.  
 XX 03-AUG-2001.  
 XX 05-FEB-2001; 2001WO-US03800.  
 XX 03-FEB-2000; 2000US-0496914.  
 XX 27-APR-2000; 2000US-0560875.  
 XX (HYSSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Drmanac RT;  
 DR WPI: 2001-457740/49.  
 DR N-PSDB; ABA09052.  
 XX Human proteins and DNA encoding sequences useful for preventing,  
 XX treating or ameliorating a medical condition in a mammalian subject  
 XX e.g. arthritis and cancer  
 XX Claim 20; Page 256-257; 1963pp; English.  
 XX Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 XX sequences ABA0223-ABA09574 represent nucleic acids encoding them. The  
 XX invention also relates to vectors and recombinant host cells comprising a  
 XX nucleotide of the invention, methods of producing the novel polypeptides,  
 XX antibodies against the polypeptides, methods of detecting the nucleotides  
 XX or polypeptides in a sample, and methods of identifying compounds which  
 XX bind to polypeptides of the invention. Although novel, many of the  
 XX polypeptides of the invention have homology to known proteins, thereby  
 XX giving an insight into their probable biological activities, and hence,  
 XX potential therapeutic applications. The polypeptides of the invention may  
 XX have various activities, including cytokine, cell proliferation or cell  
 XX differentiation activities; stem cell growth factor activity;  
 XX haematopoietic regulatory activity; tissue growth activity;  
 XX immunomodulatory activity; activin- or inhibin-related activities;  
 XX chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 XX thrombolytic activities; receptor or ligand activities; or may be  
 XX involved in oncogenesis, cancer cell proliferation or metastasis.  
 XX Depending on their biological activities, polypeptides and nucleotides of  
 XX the invention are useful for preventing, treating or ameliorating medical  
 XX conditions, e.g., by protein or gene therapy. Such conditions include  
 XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 XX proliferative retinopathy, atherosclerosis, coronary heart disease,  
 XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 XX vascular growth. Polypeptides involved with tissue regeneration and  
 XX repair (or nucleic acids encoding them) may be used to promote wound  
 XX healing (e.g., of burns, incisions and ulcers), while those with  
 XX immunomodulatory activities may be used in the treatment of viral,  
 XX bacterial and fungal infections in addition to immune disorders.  
 XX Polypeptides with growth factor activity may be used in cell cultures to  
 XX promote cell growth. For example, such polypeptides may be used to  
 XX manipulate stem cells in culture to give rise to neuroepithelial cells  
 XX that can be used to augment or replace cells damaged by illness,  
 XX autoimmune disease or accidental damage. The polypeptides and nucleotides  
 XX may also be used in the diagnosis of the above conditions, and in drug  
 XX screening techniques. The present sequence represents a novel human  
 XX polypeptide of the invention.  
 XX Sequence 1435 AA;  
 XX Query Match 100.0%; Score 136; DB 22; Length 1435;  
 XX Best Local Similarity 100.0%; Pred. No. 2.9e-12;  
 XX Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28  
 DB 217 VYVFSTEMANKAAEAVLKQGVETIVSFH 244  
 RESULT 3  
 AAO05855  
 ID AAO05855 standard; Protein; 140 AA.  
 XX AAO05855;  
 XX 06-NOV-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 19747.  
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 XX tissue growth factor; immunomodulatory; cancer; leukaemia;  
 XX nervous system disorders; arthritis; inflammation.

OS Homo sapiens.  
 XX WO200164835-A2.  
 PN  
 XX  
 XX 07-SEP-2001.  
 DD  
 XX  
 XX 26-FEB-2001; 2001WO-US04927.  
 XX  
 XX 28-FEB-2000; 2000US-0515126.  
 PR  
 XX 18-MAY-2000; 2000US-0577409.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX  
 XX Tang YT, Liu C, Dmanac RT;  
 PI  
 XX WPI; 2001-514938/56.  
 DR  
 XX N-PSDB; AA185786.  
 DR  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 PT  
 XX  
 XX Claim 20: SEQ ID NO 19747; 1399pp + Sequence Listing; English.  
 PS  
 XX The invention relates to human polynucleotides (AA179941-AA193841) and  
 CC the encoded proteins (AA000010-AA013910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 140 AA;  
 SQ  
 Query Match 54.9%; Score 129; DB 22; Length 140;  
 Best Local Similarity 92.9%; Pred. No. 2.3e-12;  
 Matches 26; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28  
 Db 59 VYVFSTEMANKAAEAVLKQGVETIVSFH 86  
 RESULT 4  
 ABB58779  
 ID ABB58779 standard; Protein; 1429 AA.  
 XX  
 XX ABB58779;  
 AC  
 XX  
 XX 26-MAR-2002 (first entry)  
 DT  
 XX  
 XX Drosophila melanogaster polypeptide SEQ ID NO 3129.  
 DE  
 XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 KW  
 XX Drosophila melanogaster.  
 OS  
 XX WO200171042-A2.  
 PN  
 XX  
 XX 27-SEP-2001.  
 PD  
 XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 PF  
 XX  
 XX 23-MAR-2000; 2000US-191637P.  
 PR  
 XX 11-JUL-2000; 2000US-0614150.  
 PR

XX (PEKE ) PE CORP NY.  
 PA  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 PI  
 XX WPI; 2001-656860/75.  
 DR  
 XX N-PSDB; ABL02882.  
 DR  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 PT  
 XX  
 XX Disclosure; SEQ ID NO 3129; 21pp + Sequence Listing; English.  
 PS  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB57737-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 XX Sequence 1429 AA;  
 SQ  
 Query Match 72.1%; Score 98; DB 22; Length 1429;  
 Best Local Similarity 57.1%; Pred. No. 2.8e-06;  
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 VYVFSTEMANKAAEAVLKQGVETIVSFH 28  
 Db 323 IFVFSTOLANKGAEVLSGGFQTIIAYH 350  
 RESULT 5  
 AAB71228  
 ID AAB71228 standard; Protein; 1464 AA.  
 XX  
 XX AAB71228;  
 AC  
 XX  
 XX 18-NOV-2002 (first entry)  
 DT  
 XX  
 XX D. melanogaster lgs protein.  
 DE  
 XX  
 XX Legless; fruitfly; lgs; Wnt/Wingless signaling pathway; Wnt; Wg;  
 KW tissue proliferation; tumour; cytoskeletal; cellular disorder; colon;  
 KW blood disorder; cancer; breast; head and neck cancer; brain; thyroid;  
 KW medulloblastoma; skin cancer; tissue regeneration; tissue repair.  
 XX  
 XX Drosophila melanogaster.  
 OS  
 XX US2002086986-A1.  
 PN  
 XX  
 XX 04-JUL-2002.  
 PD  
 XX  
 XX 27-JUL-2001; 2001US-0915543.  
 PF  
 XX  
 XX 28-JUL-2000; 2000US-221502P.  
 PR  
 XX (BASL/) BASLER K.  
 PA (BRUN/) BRUNNER E.  
 PA (FROE/) FROESCH B.  
 PA (KEAM/) KEAMES T.  
 PA (PETE/) PETER O.  
 XX  
 XX Basler K, Brunner E, Froesch B, Kramps T, Peter O;  
 PI WPI; 2002-635689/68.  
 DR N-PSDB; AAF88466.  
 DR  
 XX

degenerative



CC protection or regeneration and treatment of lung or liver fibrosis.  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage for this patent did not form part of the printed  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 114 AA;

Query Match 66.9%; Score 91; DB 23; Length 114;  
 Best Local Similarity 57.1%; Pred. No. 1.7e-06;  
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQVETIVSFH 28  
 |||||:|||||||:|||||  
 DB 78 VYVFTTHLANTAABAVLQGRADSLAYH 105

RESULT 8

AAU78461  
 ID AAU78461 standard; Protein; 320 AA.

XX AC AAU78461;

XX DT 02-JUL-2002 (first entry)

XX DE Mouse beta-catenin nuclear localised protein #2.

XX KW Mouse; beta-catenin nuclear localised protein; cancer;  
 gene therapy; EST; expressed sequence tag.

XX OS Mus musculus.

XX PN WO200224738-A1.

XX PD 28-MAR-2002.

XX PF 19-SEP-2001; 2001WO-JP08140.

XX PR 22-SEP-2000; 2000JP-0287876.

XX RA (KYOW ) KYOWA HAKKO KOGYO KK.

XX RI Akiyama T, Adachi S;

XX DR MPI; 2002-330014/36.

XX DR N-PSDB; ABK47632.

XX PT New beta-catenin nuclear localised protein for diagnosis and treatment  
 of diseases associated with nuclear localisation of beta-catenin e.g.  
 cancer

XX PS Claim 2; Page 91-92; 113pp; Japanese.

XX CC The invention relates to a beta-catenin nuclear localised protein  
 and DNA encoding the protein. The protein and encoding DNA are  
 CC applicable in diagnosis and treatment of diseases associated with  
 CC nuclear localisation of beta-catenin e.g. cancer, including gene  
 CC therapy. The present sequence represents the amino acid sequence of  
 CC mouse beta-catenin nuclear localised protein #2.

XX SQ Sequence 320 AA;

Query Match 49.3%; Score 67; DB 23; Length 320;  
 Best Local Similarity 57.1%; Pred. No. 0.035;  
 Matches 12; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKQVETIVSFH 28

DB 1 LANTAABAVLQGRADSLAYH 21

RESULT 9

RAO07544  
 ID RAO07544 standard; Protein; 113 AA.

XX AC RAO07544;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 21436.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR MPI; 2001-514838/56.

XX DR N-PSDB; AAI87475.

XX PT Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders

XX PS Claim 20; SEQ ID NO 21436; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93941) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.

XX CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 113 AA;

Query Match 47.8%; Score 65; DB 22; Length 113;  
 Best Local Similarity 75.0%; Pred. No. 0.021;  
 Matches 15; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKQVETIVSF 27

DB 81 MCVQAAEAVLKQVETDASF 100

RESULT 10

AAAG33446  
 ID AAG33446 standard; Protein; 360 AA.

XX AC AAG33446;

XX DT 18-OCT-2000 (first entry)

XX DE Zea mays protein fragment SEQ ID NO: 40525.

XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence; corn.  
XX Zea mays subsp. mays.  
XX EF1033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-0301439.  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129445.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131449.  
XX 04-MAY-1999; 99US-0132048.  
XX 06-MAY-1999; 99US-0132407.  
XX 05-MAY-1999; 99US-0132484.  
XX 06-MAY-1999; 99US-0132485.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 14-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 18-MAY-1999; 99US-0134370.  
XX 19-MAY-1999; 99US-0134768.  
XX 20-MAY-1999; 99US-0134941.  
XX 21-MAY-1999; 99US-0135124.  
XX 24-MAY-1999; 99US-0135353.  
XX 25-MAY-1999; 99US-0135429.  
XX 27-MAY-1999; 99US-0136021.  
XX 28-MAY-1999; 99US-0136392.  
XX 01-JUN-1999; 99US-0136782.  
XX 03-JUN-1999; 99US-0137222.  
XX 04-JUN-1999; 99US-0137528.  
XX 07-JUN-1999; 99US-0137502.  
XX 08-JUN-1999; 99US-0137724.  
XX 10-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 14-JUN-1999; 99US-0138847.  
XX 16-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 17-JUN-1999; 99US-0139453.  
XX 18-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.  
XX 18-JUN-1999; 99US-0139460.  
XX 18-JUN-1999; 99US-0139461.  
XX 18-JUN-1999; 99US-0139462.  
XX 18-JUN-1999; 99US-0139463.  
XX 18-JUN-1999; 99US-0139750.  
XX 18-JUN-1999; 99US-0139763.  
XX 21-JUN-1999; 99US-0139817.  
XX 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142154.  
PR 06-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143242.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147304.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0148368.  
PR 17-AUG-1999; 99US-0148175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.

```

Query Match      38.2%;   Score 52; DB 21; Length 360;
Best Local Similarity 50.0%;   Pred. NO. 9.4;
Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY      4 FSTEMANKAAERVLKQVETIVSF 27
      |||:|||
      |||:|||
Db       97 FSTRLLANNLENLVKEGPETIAAF 120
      |||:|||
      |||:|||

RESULT 11
AAG33445
ID AAG33445 standard; Protein; 448 AA.
XX
AC AAG33445;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 40524.
XX
DE Zea mays subsp. mays.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence; corn.
XX
OS
OS Zea mays subsp. mays.
XX
XX

```

PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 06-JUL-1999; 99US-0142380.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142820.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 19-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 23-JUL-1999; 99US-0145132.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 26-JUL-1999; 99US-0145224.  
 PR 27-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159323.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161362.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 52; DB 21; Length 448;

Best Local Similarity 50.0%; Pred. No. 12;

Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKGQVETIVSF 27  
 ||| : ||| ||| ||| : |||  
 DB 185 FSTRLANNLENLVLEKGPETIAAP 208

RESULT 12

AAG33444

ID AAG33444 standard; Protein; 509 AA.

XX AAG33444;

XX AAG33444;

DT 18-OCT-2000 (first entry)

DE Zea mays protein fragment SEQ ID NO: 40523.

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence; corn.

OS Zea mays subsp. mays.

PN EP1033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.



PR 05-OCT-1999; 99US-015753.  
 PR 06-OCT-1999; 99US-0157853.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159337.  
 PR 14-OCT-1999; 99US-0159338.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 52; DB 21; Length 509;  
 Best Local Similarity 50.0%; Pred. No. 14;  
 Matches 12; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQVETIVSF 27  
 ||| :|| ||| ||| :||  
 Db 246 FSTRLLNNLENLKKGPTIAAF 269

RESULT 13  
 ABB60387  
 ID ABB60387 standard; Protein; 1049 AA.

XX AC ABB60387;  
 XX AC ABB60387;  
 DT 26-MAR-2002 (first entry)  
 DE Drosophila melanogaster polypeptide SEQ ID NO 7953.  
 XX XX  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 XX XX pharmaceutical.  
 XX XX  
 XX Drosophila melanogaster.  
 XX XX  
 XX WO200171042-A2.  
 XX XX  
 XX 27-SEP-2001.  
 XX XX  
 XX 23-MAR-2001; 2001WO-US09231.  
 XX XX  
 XX 23-MAR-2000; 2000US-191637P.  
 XX XX  
 XX 11-JUL-2000; 2000US-061450.  
 XX XX  
 XX (PEKE ) PE CORP NY.  
 XX XX  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 XX XX  
 XX WPI; 2001-656860/75.  
 DR DR  
 XX N-PSDB; ABL04490.

XX XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 interactions -  
 XX XX  
 XX PS Disclosure; SEQ ID NO 7953; 21pp + Sequence Listing; English.  
 XX XX  
 XX CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABE57737-ABE72072).  
 CC CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX XX  
 XX SQ Sequence 1049 AA;  
 Query Match 37.5%; Score 51; DB 22; Length 1049;  
 Best Local Similarity 40.7%; Pred. No. 49;  
 Matches 11; Conservative 5; Mismatches 11; Indels 0; Gaps 0;  
 QY 2 YVFSTEMANKAAEAVLKQVETIVSFH 28  
 ||| :|| ||| ||| :||  
 Db 471 YVSCSDMAASATEAVRSGLKIIPHH 497  
 RESULT 14  
 ABB89793  
 ID ABB89793 standard; Protein; 130 AA.  
 XX XX  
 XX AC ABB89793;  
 XX AC ABB89793;  
 DT 24-MAY-2002 (first entry)  
 XX XX  
 XX DE Human polypeptide SEQ ID NO 2169.  
 XX XX  
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
 XX antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
 XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
 XX cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;  
 XX neurological disease; infection; human; secreted protein.  
 XX XX  
 XX OS Homo sapiens.  
 XX XX  
 XX WO200190304-A2.  
 XX XX  
 XX 29-NOV-2001.  
 XX XX  
 XX PF 18-MAY-2001; 2001WO-US16450.  
 XX XX  
 XX PR 19-MAY-2000; 2000US-205515P.  
 XX XX  
 XX XX (HUMA-) HUMAN GENOME SCI INC.  
 XX XX  
 XX BIIRSE CE, ROSEN CA;  
 XX XX  
 XX WPI; 2002-122018/15.  
 DR DR  
 DR N-PSDB; ABL90202.  
 XX XX  
 XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
 PT prevention of neural, immune system, muscular, reproductive,  
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
 PT disorders -  
 XX XX  
 XX PS Claim 11; SEQ ID NO 2169; 2081pp + Sequence Listing; English.  
 XX XX  
 XX CC The invention relates to novel genes (ABL89449-ABL90853) and proteins  
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are

quantitating a polypeptide in tissue, as molecular weight markers and a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II) (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [www.int/pub/published/pct/sequences](http://www.int/pub/published/pct/sequences).

sq	Sequence	365 AA:
Query Match	36.6%	Score 50; DB 22; Length 365;
Best Local Similarity	29.6%	Pred. No. 20;
Matches	8	Conservative 8; Mismatches 11; Indels 0; Gaps 0;

67 YVFAYPVCNASAKTIISGLTECLIHCH 93

DATE : 14.12.2023

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	47	32.6	2639	3	US-09-080-383-3	Sequence 3, Appli
2	44	34.4	718	3	US-09-090-008-2	Sequence 2, Appli
3	44	32.4	718	3	US-09-047-453-2	Sequence 2, Appli
4	43	31.6	145	4	US-09-091-725-37	Sequence 37, Appli
5	43	31.6	438	3	US-08-486-099-93	Sequence 93, Appli
6	43	31.6	438	3	US-08-360-107A-103	Sequence 103, App
7	43	31.6	438	3	US-08-484-2238-93	Sequence 93, Appli
8	43	31.6	438	3	US-08-919-597-93	Sequence 93, Appli
9	43	31.6	438	3	US-08-475-668A-93	Sequence 93, Appli
10	43	31.6	438	3	US-08-485-551A-93	Sequence 93, Appli
11	43	31.6	438	3	US-08-471-913A-93	Sequence 93, Appli
12	43	31.6	438	3	US-08-485-2648-93	Sequence 93, Appli
13	43	31.6	438	3	US-08-474-349A-93	Sequence 93, Appli
14	43	31.6	438	4	US-08-255-208A-29	Sequence 29, Appli
15	43	31.6	438	4	US-08-470-696-93	Sequence 93, Appli
16	43	31.6	428	4	US-08-485-546B-93	Sequence 93, Appli
17	43	31.6	662	1	US-08-224-557-88	Sequence 88, Appli
18	43	31.6	662	4	US-09-354-138-88	Sequence 88, Appli
19	42	30.9	154	4	US-09-732-210-103	Sequence 103, App
20	42	30.9	154	4	US-09-732-210-210	Sequence 210, App
21	42	30.9	188	4	US-09-252-991A-29737	Sequence 29737, A
22	42	30.9	300	4	US-09-634-238-219	Sequence 219, App
23	42	30.9	358	4	US-09-934-901-88	Sequence 8, Appli
24	42	30.9	541	4	US-09-627-376-11	Sequence 11, Appli
25	42	30.9	1820	3	US-07-998-389B-8	Sequence 8, Appli
26	41.5	30.5	111	4	US-09-775-932-20	Sequence 20, Appli
27	41.5	30.5	178	1	US-08-844-085-2	Sequence 2, Appli





100

APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TRANSMISSION  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 245

CORRESPONDENCE ADDRESS: 245  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/484,223B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 438 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
MOLECULE TYPE: protein

US-08-484-223B-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STENANKAAAEAVLKQVETIVS 26  
DB 41 SLEQSNKAIEIREATQETVIA 62

## RESULT 8

US-08-919-597-93  
Sequence 93, Application US/08919597  
Patent No. 6054285

GENERAL INFORMATION:

APPLICANT: Boicognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION  
OF MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING HIV  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 273

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/919,597  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/470,896  
FILING DATE: 06-JUN-1995  
ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 438 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown

MOLECULE TYPE: protein  
US-08-919-597-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STENANKAAAEAVLKQVETIVS 26  
DB 41 SLEQSNKAIEIREATQETVIA 62

## RESULT 9

US-08-475-668A-93  
Sequence 93, Application US/08475668A  
Patent No. 6060065

GENERAL INFORMATION:

APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
TITLE OF INVENTION: TRANSMISSION

NUMBER OF SEQUENCES: 211  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/475,668A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742

REFERENCE/DOCKET NUMBER: 7872-026

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864

INFORMATION FOR SEQ ID NO: 93:

SEQUENCE CHARACTERISTICS:

LENGTH: 438 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-475-668A-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26  
DB 41 SLEQSNKAIEIRATQETVIA 62

RESULT 10  
US-08-485-551A-93  
Sequence 93, Application US/0848551A  
Patent No. 6068973

GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: METHODS FOR INHIBITION OF MEMBRANE  
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING INFLUENZA VIRUS  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 211  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/485.551A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-023  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 438 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown

MOLECULE TYPE: protein  
US-08-485-551A-93  
Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26  
DB 41 SLEQSNKAIEIRATQETVIA 62

RESULT 11  
US-08-471-913A-93  
Sequence 93, Application US/08471913A  
Patent No. 6093794

GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE  
TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING EPSTEIN-BARR VIRUS  
TITLE OF INVENTION: TRANSMISSION  
NUMBER OF SEQUENCES: 214  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/471.913A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7872-030  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 438 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-471-913A-93

Query Match 31.6%; Score 43; DB 3; Length 438;  
Best Local Similarity 36.4%; Pred. No. 63;  
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 5 STEMANKAAEAVLKGQVETIVS 26  
DB 41 SLEQSNKAIEIRATQETVIA 62

RESULT 12  
US-08-485-264A-93  
Sequence 93, Application US/08485264A  
Patent No. 6228983

GENERAL INFORMATION:  
APPLICANT: Bolognesi, Dani P.  
APPLICANT: Matthews, Thomas J.  
APPLICANT: Wild, Carl T.  
APPLICANT: Barney, Shawn O.  
APPLICANT: Lambert, Dennis M.  
APPLICANT: Petteway, Stephen R.  
APPLICANT: Langlois, Alphonse J.  
TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF  
TITLE OF INVENTION: MEMBRANE FUSION-ASSOCIATED EVENTS, INCLUDING

```

; TITLE OF INVENTION: RESPIRATORY SYNCYTIAL VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 232
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485.264A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-485-264A-93
;
; Query Match 31.6%; Score 43; DB 3; Length 438;
; Best Local Similarity 36.4%; Pred. No. 63;
; Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
;
; QY 5 STEMANKAAEAVLKQGVETIVS 26
; DB 41 SLEQSNKAIEIREATQETVIA 62
;
; RESULT 13
; US-08-474-349A-93
; Sequence 93 Application US/08/474349A
; Patent No. 6333395
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl I.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway Jr., Stephen R.
; TITLE OF INVENTION: COMPOSITIONS FOR INHIBITION OF MEMBRANE
; FUSION-ASSOCIATED EVENTS, INCLUDING HUMAN PARAINFLUENZA
; TITLE OF INVENTION: VIRUS TRANSMISSION
; NUMBER OF SEQUENCES: 517
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/474,349A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 93:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-474-349A-93
;
; Query Match 31.6%; Score 43; DB 4; Length 438;
; Best Local Similarity 36.4%; Pred. No. 63;
; Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
;
; QY 5 STEMANKAAEAVLKQGVETIVS 26
; DB 41 SLEQSNKAIEIREATQETVIA 62
;
; RESULT 14
; US-08-255-208A-29
; Sequence 29 Application US/08255208A
; Patent No. 6440656
; GENERAL INFORMATION:
; APPLICANT: Bolognesi, Dani P.
; APPLICANT: Matthews, Thomas J.
; APPLICANT: Wild, Carl T.
; APPLICANT: Barney, Shawn O.
; APPLICANT: Lambert, Dennis M.
; APPLICANT: Petteway Jr., Stephen R.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE INHIBITORS OF HIV
; TITLE OF INVENTION: TRANSMISSION
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/255,208A
; FILING DATE: 07-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7872-010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:27:44 ; Search time 7.51751 Seconds  
(without alignments)  
679.968 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136  
Sequence: 1 VYFSTEMAKAAEAVLKQVETIVSFH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	28	15	US-10-322-579-3
2	136	100.0	1426	15	US-10-322-579-15
3	98	72.1	28	15	US-10-322-579-2
4	49	36.0	984	15	US-10-128-714-3547
5	49	36.0	1058	15	US-10-128-714-8547
6	47	34.6	732	14	US-10-008-355-6
7	46	33.8	471	12	US-10-167-547C-8
8	45	33.1	507	12	US-10-167-547C-10
9	45	33.1	747	15	US-10-156-761-10721
10	44.5	32.7	170	12	US-10-342-224-92
11	44	32.4	171	15	US-10-156-761-9002
12	44	32.4	484	9	US-09-815-242-10457
13	44	32.4	718	14	US-10-117-846-2
14	43	31.6	80	11	US-09-764-891-5144
15	43	31.6	89	12	US-10-029-386-28376

16	43	31.6	134	12	US-10-029-386-31669
17	43	31.6	315	12	US-10-306-762-212
18	43	31.6	602	12	US-10-220-381-7
19	43	31.6	623	12	US-10-032-585-7819
20	43	31.6	662	11	US-09-951-061A-141
21	42	30.9	248	12	US-10-097-111-304
22	42	30.9	277	15	US-10-156-761-8036
23	42	30.9	326	15	US-10-156-761-8036
24	42	30.9	358	10	US-09-934-901-8
25	42	30.9	358	12	US-09-934-868-18
26	42	30.9	358	12	US-10-320-924-8
27	42	30.9	358	16	US-10-320-924-8
28	42	30.9	380	12	US-10-183-708-42
29	42	30.9	380	12	US-09-532-227-42
30	42	30.9	399	12	US-10-234-971-5
31	41.5	30.5	541	14	US-10-041-676A-11
32	41.5	30.5	111	10	US-09-775-952-20
33	41.5	30.5	469	8	US-08-781-986A-5245
34	41	30.1	227	9	US-09-815-242-11197
35	41	30.1	279	9	US-08-861-451A-28
36	41	30.1	370	10	US-09-712-363-187
37	41	30.1	370	12	US-10-361-460-10
38	41	30.1	451	11	US-09-910-186A-12
39	41	30.1	456	9	US-09-815-242-4991
40	41	30.1	456	9	US-09-815-242-10636
41	41	30.1	516	10	US-09-738-626-5086
42	41	30.1	568	15	US-10-156-761-11478
43	41	30.1	638	15	US-10-234-432-59
44	41	30.1	689	14	US-10-052-664-1
45	41	30.1	689	15	US-10-097-340-221

ALIGNMENTS

RESULT 1

US-10-322-579-3

Sequence 3, Application US/10322579

Publication, No. US2003011413A1

GENERAL INFORMATION:

APPLICANT: BASLER, Konrad

APPLICANT: BROEDER, Erich

APPLICANT: FROESCH, Barbara

APPLICANT: KRAMER, Thomas

APPLICANT: PETER, Oliver

TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY

TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON

FILE REFERENCE: 060361

CURRENT APPLICATION NUMBER: US/10322579

CURRENT FILING DATE: 2002-12-19

PRIOR APPLICATION NUMBER: US/09/915,543

PRIOR FILING DATE: 2001-07-27

PRIOR APPLICATION NUMBER: 60/221,502

PRIOR FILING DATE: 2000-07-28

NUMBER OF SEQ ID NOS: 22

SOFTWARE: Patent in version 3.1

SEQ ID NO 3

LENGTH: 28

TYPE: PRT

ORGANISM: Human lgs/bc19

US-10-322-579-3

Query Match 100.0%; Score 136; DB 15; Length 28;

Best Local Similarity 100.0%; Pred. No. 1e-14;

Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VYFSTEMAKAAEAVLKQVETIVSFH 28

Db 1 VYFSTEMAKAAEAVLKQVETIVSFH 28

RESULT 2

US-10-322-579-15

Sequence 15, Application US/10322579  
Publication No. US20030114413A1  
GENERAL INFORMATION:  
APPLICANT: BASLER, Konrad  
APPLICANT: BRUNNER, Erich  
APPLICANT: FROESCH, Barbara  
APPLICANT: KRAMES, Thomas  
APPLICANT: PETER, Oliver  
TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
FILE REFERENCE: Q60361  
CURRENT APPLICATION NUMBER: US/10/322.579  
CURRENT FILING DATE: 2002-12-19  
PRIOR APPLICATION NUMBER: US/09/915.543  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: 60/421.502  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 15  
LENGTH: 1426  
TYPE: PRT  
ORGANISM: Human lgs/bcl9  
US-10-322-579-15

Query Match 100.0%; Score 136; DB 15; Length 1426;  
Best Local Similarity 100.0%; Pred. No. 1.1e-12;  
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VVVFSTEMANKAAEAVLKGQVETIVSFH 28  
DB 177 VVVFSTEMANKAAEAVLKGQVETIVSFH 204

## RESULT 3

US-10-322-579-2  
Sequence 2, Application US/10322579  
Publication No. US20030114413A1  
GENERAL INFORMATION:  
APPLICANT: BASLER, Konrad  
APPLICANT: BRUNNER, Erich  
APPLICANT: FROESCH, Barbara  
APPLICANT: KRAMES, Thomas  
APPLICANT: PETER, Oliver  
TITLE OF INVENTION: ESSENTIAL DOWNSTREAM COMPONENT OF THE WINGLESS SIGNALING PATHWAY  
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC APPLICATIONS BASED THEREON  
FILE REFERENCE: Q60361  
CURRENT APPLICATION NUMBER: US/10/322.579  
CURRENT FILING DATE: 2002-12-19  
PRIOR APPLICATION NUMBER: US/09/915.543  
PRIOR FILING DATE: 2001-07-27  
PRIOR APPLICATION NUMBER: 60/421.502  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 22  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 28  
TYPE: PRT  
ORGANISM: Drosophila lgs  
US-10-322-579-2

Query Match 72.1%; Score 98; DB 15; Length 28;  
Best Local Similarity 57.1%; Pred. No. 1.1e-08;  
Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 VVVFSTEMANKAAEAVLKGQVETIVSFH 28  
DB 1 IFVSTQLANKAESVLSGGQFIIAYH 28

## RESULT 4

US-10-128-714-3547  
Sequence 3547, Application US/10128714

Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroshkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
TITLE OF INVENTION: Methods of Use  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128.714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285.697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287.066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295.890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303.899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316.362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3547  
LENGTH: 984  
TYPE: PRT  
ORGANISM: Aspergillus fumigatus  
US-10-128-714-3547

Query Match 36.0%; Score 49; DB 15; Length 984;  
Best Local Similarity 52.6%; Pred. No. 52;  
Matches 10; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKGQV 21  
DB 500 VYRANWANKAAAVLKS 518

## RESULT 5

US-10-128-714-8547  
Sequence 8547, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jiang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Eroshkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and  
TITLE OF INVENTION: Methods of Use  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128.714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285.697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287.066  
PRIOR FILING DATE: 2001-04-27  
PRIOR APPLICATION NUMBER: US 60/295.890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303.899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316.362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8547  
LENGTH: 1058  
TYPE: PRT  
ORGANISM: Aspergillus fumigatus  
US-10-128-714-8547



Query Match 36.0%; Score 49; DB 15; Length 1058;  
Best Local Similarity 52.6%; Pred. No. 57;  
Matches 10; Conservative 5; Mismatches 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKQVET 21  
DB 500 VYFANWANKAAEAVLKSKL 518

## RESULT 6

US-10-008-355-6  
; Sequence 6, Application US/10008355  
; Publication No. US20020164759A1  
; GENERAL INFORMATION:  
; APPLICANT: Travis, James  
; APPLICANT: Potempa, Jan S  
; APPLICANT: Banbula, Agnieszka  
; TITLE OF INVENTION: Dipeptidylpeptidases And Methods Of Use  
; FILE REFERENCE: 235, 00440101  
; CURRENT APPLICATION NUMBER: US/10/008,355  
; PRIOR FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: US 60/246,827  
; PRIOR FILING DATE: 2000-11-08  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 6  
; LENGTH: 732  
; TYPE: PRT  
; ORGANISM: Shewanella putrefaciens  
US-10-008-355-6

Query Match 34.6%; Score 47; DB 14; Length 732;  
Best Local Similarity 39.1%; Pred. No. 76;  
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKQVET 23  
DB 517 VALYDYNWAEKAEKILAGKLT 539

## RESULT 7

US-10-167-547C-8  
; Sequence 8, Application US/10167547C  
; Publication No. US20030170653A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. du Pont de Nemours and Company  
; APPLICANT: Damude, Howard G.  
; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gamma  
; FILE REFERENCE: CL1804 US NA  
; CURRENT APPLICATION NUMBER: US/10/167,547C  
; PRIOR FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: 60/297198  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: Microsoft Office 07  
; SEQ ID NO 8  
; LENGTH: 471  
; TYPE: PRT  
; ORGANISM: tulip pistol  
US-10-167-547C-8

Query Match 33.8%; Score 46; DB 12; Length 471;  
Best Local Similarity 41.7%; Pred. No. 64;  
Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQVETIVSF 27  
DB 204 FSTRLANNLEKILTEGPETIAAF 227

## RESULT 8

; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress

US-10-167-547C-10  
; Sequence 10, Application US/10167547C  
; Publication No. US20030170653A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. du Pont de Nemours and Company  
; APPLICANT: Damude, Howard G.

; TITLE OF INVENTION: A Biological Method for the Production of Alpha-Methylene-Gamma  
; FILE REFERENCE: CL1804 US NA  
; CURRENT APPLICATION NUMBER: US/10/167,547C  
; PRIOR FILING DATE: 2003-03-17  
; PRIOR APPLICATION NUMBER: 60/297198  
; PRIOR FILING DATE: 2001-06-08  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: Microsoft Office 07  
; SEQ ID NO 10  
; LENGTH: 507  
; TYPE: PRT  
; ORGANISM: Alstroemeria  
US-10-167-547C-10

Query Match 33.1%; Score 45; DB 12; Length 507;  
Best Local Similarity 37.5%; Pred. No. 1e+02;  
Matches 9; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 4 FSTEMANKAAEAVLKQVETIVSF 27  
DB 246 FSTRLANNLEKILTEGPETIAAF 269

## RESULT 9

US-10-156-761-10721  
; Sequence 10721, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHISA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-23  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10721  
; LENGTH: 747  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10721

Query Match 33.1%; Score 45; DB 15; Length 747;  
Best Local Similarity 39.1%; Pred. No. 1.6e+02;  
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAEAVLKQVETIV 25  
DB 250 VFATSSLNRAQALDRGETERAV 272

## RESULT 10

US-10-342-224-92  
; Sequence 92, Application US/10342224  
; Publication No. US20030162294A1  
; GENERAL INFORMATION:  
; APPLICANT: Nathalie Verbruggen  
; TITLE OF INVENTION: Genes Involved in Tolerance to Environmental Stress

FILE REFERENCE: CNN-012US  
 CURRENT APPLICATION NUMBER: US/10/342,224  
 CURRENT FILING DATE: 2003-01-13  
 PRIOR APPLICATION NUMBER: US/09/762,154  
 PRIOR FILING DATE: 2002-02-02  
 PRIOR APPLICATION NUMBER: EP 98202634.6  
 PRIOR FILING DATE: 1998-08-04  
 NUMBER OF SEQ ID NOS: 123  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 92  
 LENGTH: 170  
 TYPE: PRT  
 ORGANISM: Arabidopsis thaliana  
 US-10-342-224-92

Query Match 32.7%; Score 44.5; DB 12; Length 170;  
 Best Local Similarity 52.0%; Pred. No. 33;  
 Matches 13; Conservative 0; Mismatches 7; Indels 5; Gaps 1;

QY 6 TEMANKAAEAVL-----KGQVETIV 25  
 DB 85 TEIVKTVVLLDTLEKKQVETIV 109

RESULT 11  
 US-10-156-761-9002  
 Sequence 9002, Application US/10156761  
 Publication No. US20030119018A1  
 GENERAL INFORMATION:  
 APPLICANT: OMURA, SATOSHI  
 APPLICANT: IKEDA, HARUO  
 APPLICANT: ISHIKAWA, JUN  
 APPLICANT: HIRAKAWA, HIROSHI  
 APPLICANT: SHIEA, TADAYOSHI  
 APPLICANT: SAKAKI, YOSHIOYUKI  
 APPLICANT: HATTORI, MASAHIRA  
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
 FILE REFERENCE: 249-462  
 CURRENT APPLICATION NUMBER: US/10/156,761  
 CURRENT FILING DATE: 2002-05-23  
 PRIOR APPLICATION NUMBER: JP 2001-204089  
 PRIOR FILING DATE: 2001-05-30  
 PRIOR APPLICATION NUMBER: JP 2001-272697  
 PRIOR FILING DATE: 2001-08-02  
 NUMBER OF SEQ ID NOS: 15109  
 SEQ ID NO 9002  
 LENGTH: 171  
 TYPE: PRT  
 ORGANISM: Streptomyces avermitilis  
 US-10-156-761-9002

Query Match 32.4%; Score 44; DB 15; Length 171;  
 Best Local Similarity 35.7%; Pred. No. 40;  
 Matches 10; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 VYVFSTEMANKAAEAVLKGQVETIVSFH 28  
 DB 29 IFCYRSMNAKAAEQVLVDQWRGILAVH 56

RESULT 12  
 US-09-815-242-10457  
 Sequence 10457, Application US/09815242  
 Patent No. US20020061569A1  
 GENERAL INFORMATION:  
 APPLICANT: Haseibeck, Robert  
 APPLICANT: Ohlsen, Kari L.  
 APPLICANT: Zyskind, Judith W.  
 APPLICANT: Wall, Daniel  
 APPLICANT: Trawick, John D.  
 APPLICANT: Carr, Grant J.  
 APPLICANT: Yamamoto, Robert T.  
 APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in  
 FILE REFERENCE: Prokaryotes  
 FILE REFERENCE: ELITRA-011A  
 CURRENT APPLICATION NUMBER: US/09/815,242  
 CURRENT FILING DATE: 2003-03-21  
 PRIOR APPLICATION NUMBER: 60/191,078  
 PRIOR FILING DATE: 2000-03-21  
 PRIOR APPLICATION NUMBER: 60/206,848  
 PRIOR FILING DATE: 2000-05-23  
 PRIOR APPLICATION NUMBER: 60/207,727  
 PRIOR FILING DATE: 2000-05-26  
 PRIOR APPLICATION NUMBER: 60/242,578  
 PRIOR FILING DATE: 2000-10-23  
 PRIOR APPLICATION NUMBER: 60/253,625  
 PRIOR FILING DATE: 2000-11-27  
 PRIOR APPLICATION NUMBER: 60/257,931  
 PRIOR FILING DATE: 2000-12-22  
 PRIOR APPLICATION NUMBER: 60/269,308  
 PRIOR FILING DATE: 2001-02-16  
 NUMBER OF SEQ ID NOS: 14110  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 10457  
 LENGTH: 484  
 TYPE: PRT  
 ORGANISM: Escherichia coli  
 US-09-815-242-10457

Query Match 32.4%; Score 44; DB 9; Length 484;  
 Best Local Similarity 30.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 8 MANKAAEAVLKGQVETIVSF 27  
 DB 52 LLKRSVSVILKGKTIIGF 71

RESULT 13  
 US-10-117-846-2  
 Sequence 2, Application US/10117846  
 Publication No. US20020186673A1  
 GENERAL INFORMATION:  
 APPLICANT: Fuller, Margaret T  
 APPLICANT: Hales, Karen G.  
 APPLICANT: Santel, Ansgar H.  
 TITLE OF INVENTION: Mitofusins, Fzo Homologs and Functional  
 FILE REFERENCE: STAN-063CIEP  
 CURRENT APPLICATION NUMBER: US/10/117,846  
 CURRENT FILING DATE: 2002-04-05  
 PRIOR APPLICATION NUMBER: 09/413,285  
 PRIOR FILING DATE: 1999-10-06  
 PRIOR APPLICATION NUMBER: PCT/US00/27871  
 PRIOR FILING DATE: 2000-10-06  
 NUMBER OF SEQ ID NOS: 46  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 718  
 TYPE: PRT  
 ORGANISM: Drosophila melanogaster  
 US-10-117-846-2

Query Match 32.4%; Score 44; DB 14; Length 718;  
 Best Local Similarity 34.6%; Pred. No. 2.2e+02;  
 Matches 9; Conservative 5; Mismatches 10; Indels 2; Gaps 1;

QY 2 YVFS--TEMANKAAEAVLKGQVETIV 25  
 DB 403 YCFEELTENTORVGRCVLSQIKTLI 428

RESULT 14  
 US-09-764-891-5144  
 Sequence 5144, Application US/09764891

```

; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5144
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-891-5144

Query Match      31.6%; Score 43; DB 11; Length 80;
Best Local Similarity 34.8%; Pred. No. 23;
Matches      8; Conservative      7; Mismatches      8; Indels      0; Gaps      0;

Qy      1  VYVFSTEMANKAAEAVLKQGVET 23
Db      46  MHIFSEYANKPVTATSAHMQT 68

RESULT 15
US-10-029-386-28376
; Sequence 28376, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David K.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 14288
; SOFTWARE: Annotax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28376
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR4.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.92
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.72
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P03364, EVALUE 5.00e-03
US-10-029-386-28376

Query Match      31.6%; Score 43; DB 12; Length 89;
Best Local Similarity 29.6%; Pred. No. 26;
Matches      8; Conservative      6; Mismatches      13; Indels      0; Gaps      0;

Qy      2  VYVFSTEMANKAAEAVLKQGVETVSPH 28
Db      21  YVFAYPAHNSAKTTIHGLTECLHCH 47

Search completed: November 13, 2003, 09:39:45
Job time : 7.51751 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 : Search time 3.81323 Seconds  
(without alignments)  
706.153 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136  
Sequence: 1 VVVFSTEMANKAEAVLKQGVETIVSF 28

Scoring table: BLOSUM62

Gap: 10.0, Gapext 0.5

Searched: 283308 seqs, 9616862 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:\*\*

1: pir1:\*\*

2: pir2:\*\*

3: pir3:\*\*

4: pir4:\*\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	44.1	205	I40812	porphobilinogen sy
2	55	37.4	234	T27465	hypothetical prote
3	51	37.5	240	G97070	Zn-dependent hydro
4	49	36.0	218	D83398	protein F1L21.2 [
5	49	36.0	538	F71339	probable response
6	49	36.0	586	D83315	Sgat protein VCA02
7	48.5	35.7	311	C81484	probable D-2-hydro
8	48	35.3	330	E84074	dihydroxyacetone k
9	47	34.6	319	T01822	hypothetical prote
10	47	34.6	352	F90179	protease (Xaa-pro
11	47	34.6	662	D06233	probable bacteriop
12	46.5	34.2	586	T29625	hypothetical prote
13	46	33.8	330	S08500	QUTG protein - Eme
14	46	33.8	363	C92962	branched-chain fat
15	46	33.8	461	E95887	probable aminotran
16	46	33.8	1289	F72308	hypothetical prote
17	46	33.8	2123	S55089	probable acetyl-Co
18	45.5	33.5	256	S04363	class II histocomp
19	45.5	33.5	268	A9261	glutamate CoA-tr
20	45.5	33.5	504	S54744	cellulase (EC 3.2.
21	45.5	33.5	505	T32478	endoglucanase - Er
22	45	33.1	131	H27465	hypothetical prote
23	45	33.1	451	B96495	hypothetical prote
24	45	33.1	555	H96762	hypothetical prote
25	44.5	32.7	48	D90907	hypothetical prote
26	44.5	32.7	50	F85710	unknown protein en
27	44	32.4	126	T43131	hypothetical prote
28	44	32.4	265	T14645	hypothetical prote
29	44	32.4	318	C64445	conserved hypothet

30	44	32.4	340	2	T19105	phosphate carrier
31	44	32.4	350	2	A85056	probable transposo
32	44	32.4	395	2	AH3455	acriflavin resista
33	44	32.4	437	2	A72498	probable DNA/panto
34	44	32.4	476	2	AG1051	probable transport
35	44	32.4	484	2	A86116	hypothetical prote
36	44	32.4	484	2	D65230	hypothetical 52.9
37	44	32.4	484	2	A98275	hypothetical prote
38	44	32.4	533	2	T05092	probable 1-2-diacy
39	44	32.4	609	2	JCS555	vibriolysin (EC 3.
40	44	32.4	618	2	G72709	probable DNA ligase
41	44	32.4	980	2	T39630	valine-tRNA ligase
42	44	32.4	4427	2	FN0637	polyketide synthas
43	43	31.6	119	2	T18644	hypothetical prote
44	43	31.6	159	2	T40440	6,7-dimethyl-8-rib
45	43	31.6	234	2	AD3133	hypothetical prote

ALIGNMENTS

RESULT 1

I40812

porphobilinogen synthase (EC 4.2.1.24) - Clostridium josui (fragment)

N/Alternate names: delta-aminolevulinic acid dehydratase

C/Species: Clostridium josui

C/Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 16-Jul-1999

C/Accession: I40812

R/Fujino, E.; Fujino, T.; Karita, S.; Sakka, K.; Ohmiya, K.

J. Bacteriol. 177, 5169-5175, 1995

A/Title: Cloning and sequencing of some genes responsible for porphyrin biosynthesis

A/Reference number: A57344; MUID:95394829; PMID:7665501

A/Accession: I40812

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-205 <RES>

A/Cross-references: GB:D28503; NID:9536874; PIDN:BAA05863.1; PID:G556484

C/Genetics:

A/Gene: hemb

C/Superfamily: porphobilinogen synthase

C/Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 44.1%; Score 60; DB 2; Length 205;  
Best Local Similarity 46.2%; Pred No. 0.23; Indels 0; Gaps 0;  
Matches 12; Conservative 5; Mismatches 9;

QY 2 VVVFSTEMANKAEAVLKQGVETIVSF 27

Db 51 YHFSFDMVGKAEALAKADVRSLP 76

RESULT 2

T27465

hypothetical protein Y87G2A.m - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T27465

R/White, S.

submitted to the EMBL Data Library, September 1999

A/Reference number: Z20371

A/Accession: T27465

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-634 <WIL>

A/Cross-references: EMBL:AL110500; NID:e1542314; PIDN:CA854487.1; CESP:Y87G2A.m

A/Experimental source: clone Y87G2A

C/Genetics:

A/Gene: CESP:Y87G2A.m

A/Introns: 74/1; 270/1

Query Match 40.4%; Score 55; DB 2; Length 634;  
Best Local Similarity 44.4%; Pred No. 4.2; Indels 0; Gaps 0;  
Matches 12; Conservative 2; Mismatches 13;

\_\_\_\_\_

A:Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyp

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: CB1380

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-311 <PAR>

A:Cross-references: GB:AL119075; GB:AL111169; NID:G6967817; PIDN:CB874209.1; PID:G696784

A:Experimental source: serotype O2, strain NCIC 11158

C:Genetics:

A:Gene: Cj0373

C:Superfamily: phosphoglycerate dehydrogenase

Query Match 35.7%; Score 48.5; DB 2; Length 311;

Best Local Similarity 42.4%; Pred. No. 18;

Matches 14; Conservative 8; Mismatches 8; Indels 5; Gaps 2;

QY 1 VYFSTEMANKAAABV---LKGQVET--IVSPH 28

DB 171 IYVYSTGANKNADPVHLEKDLATCDTIISIH 203

RESULT 8

E84074

A:Title: dihydroxyacetone kinase BH3397 [imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: E84074

R:Takami, H.; Makasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hirai

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E84074

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-330 <STO>

A:Cross-references: GB:AP001518; GB:BA000004; NID:G10175792; PIDN:BAB07116.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3397

Query Match 35.3%; Score 48; DB 2; Length 330;

Best Local Similarity 42.3%; Pred. No. 23;

Matches 11; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

QY 2 VYFSTEMANKAAABVILKGQVETIVSF 27

DB 270 VYFVNDVANKLTSEGLNIQPKVGSF 295

RESULT 9

T01822

A:Title: hypothetical protein T27D20.16 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 26-Feb-1999 #sequence\_revision 26-Feb-1999 #text\_change 10-Dec-1999

C:Accession: T01822

R:Edwards, J.; Wollam, C.; Dubbelde, C.

submitted to the EMBL Data Library, August 1998

A:Description: The sequence of A. thaliana T27D20.

A:Reference number: Z1441

A:Accession: T01822

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-319 <SDW>

A:Cross-references: EMBL:AF076274; NID:G3293583; PID:G3377852

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 4

A:Introns: 49/3; 151/3; 210/3; 269/2

A:Superfamily: Arabidopsis hypothetical protein FTN22.18

Query Match 34.6%; Score 47; DB 2; Length 319;

Best Local Similarity 48.0%; Pred. No. 31;

Matches 12; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 5 STENANK--AAEAVILKGQVETIVSF 27

DB 243 STELSQKLAAAEALIANQAEKITSF 267

RESULT 10

F90179

A:Title: Prolidase (Xaa-Pro dipeptidase) [pepQ] [imported] - *Sulfolobus solfataricus*

C:Species: *Sulfolobus solfataricus*

C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001

C:Accession: F90179

R:Sheng, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Agoc, H.P.; Redd

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: *Sulfolobus solfataricus* complete genome.

A:Reference number: A99139

A:Accession: F90179

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-352 <KUR>

A:Cross-references: GB:AE006641; NID:G13813507; PIDN:AAK40693.1; GSPDB:GN00155

C:Genetics:

A:Gene: pepQ

C:Superfamily: X-Pro aminopeptidase

Query Match 34.6%; Score 47; DB 2; Length 352;

Best Local Similarity 43.5%; Pred. No. 35;

Matches 10; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 2 VYFSTEMANKAAABVILKGQVETI 24

DB 232 FVFKNSEAKYVTVLEAQMEAL 254

RESULT 11

AD0623

A:Title: Probable bacteriophage protein STY1061 [imported] - *Salmonella enterica* subsp. enteric

C:Species: *Salmonella enterica* subsp. enterica serovar typhi

A:Note: This species has also been called *Salmonella typhi*

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002

C:Accession: AD0623

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

. S.; Moule, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K

A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* ser

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AD0623

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-662 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05454.1; PID:G16502215; GSPDB:GN00176

C:Genetics:

A:Gene: STY1061

Query Match 34.6%; Score 47; DB 2; Length 662;

Best Local Similarity 40.9%; Pred. No. 67;

Matches 9; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 3 VFSTEMANKAAABVILKGQVETI 24

DB 600 IYSRELINKANVAGISGKYTEV 621

RESULT 12

T29695

A:Title: hypothetical protein T19H9.1 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C/Accession: T29695  
 R/Du, Z.; Gattung, S.  
 Submitted to the EMBL Data Library, November 1995  
 A/Description: The sequence of *C. elegans* cosmid T18H9.  
 A/Reference number: Z20666  
 A/Accession: T29695  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-586 <DUZ>  
 A/Cross-references: EMBL:U41746; PIDN:AAA83334.1; CESP:T18H9.1  
 C/Genetics:  
 A/Introns: 23/1; 105/3; 174/3; 303/3; 403/2; 430/3; 539/2

Query Match 34.2%; Score 46.5; DB 2; Length 586;  
 Best Local Similarity 40.7%; Pred. No. 70;  
 Matches 11; Conservative 7; Mismatches 8; Indels 1; Gaps 1;

Qy 1 VYVFSTEMANKAAEAVLKQVETIVSF 27  
 ||:||||:|  
 Db 492 VHRFATKL-QKHSKIFSTQFETIVSY 517  
 ||:||||:|

RESULT 13  
 S08500  
 QUTG protein - *Emricella nidulans*  
 C/Species: *Emricella nidulans*, *Aspergillus nidulans*  
 C/Date: 21-Nov-1993 #sequence\_revision 01-Dec-1995 #text\_change 24-Sep-1998  
 C/Accession: S08500; S11944  
 R/Hawkins, A.R.; Lamb, H.K.; Smith, M.; Keyte, J.W.; Roberts, C.F.  
 Mol. Gen. Genet. 214, 224-231, 1988  
 A/Title: Molecular organisation of the quinic acid utilization (QUT) gene cluster in *Asp. nidulans*  
 A/Reference number: S08498; MUID:89181521; PMID:2976880  
 A/Accession: S08500  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-130 <HAW>  
 A/Cross-references: EMBL:X13525; NID:q2398; PID:q2399  
 A/Note: The authors translated the codon GCG for residue 327 as Thr  
 R/Lamb, H.K.; Hawkins, A.R.; Smith, M.; Harvey, I.J.; Brown, J.; Turner, G.; Roberts, C.  
 Mol. Gen. Genet. 223, 17-23, 1990  
 A/Title: Spatial and biological characterisation of the complete quinic acid utilisation  
 A/Reference number: S11944; MUID:91080861; PMID:2175387  
 A/Accession: S11944  
 A/Molecule type: DNA  
 A/Residues: 1-326, 1 <LAM>  
 A/Experimental source: strain 2035  
 C/Genetics:  
 A/Introns: 57/1; 104/1; 250/2; 293/2

Query Match 33.8%; Score 46; DB 2; Length 330;  
 Best Local Similarity 38.9%; Pred. No. 45;  
 Matches 7; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

Qy 1 VYVFSTEMANKAAEAVLK 18  
 ||:||||:|  
 Db 13 IYAFATDLARKAGQLLLE 30  
 ||:||||:|

RESULT 14  
 C69562  
 branched-chain fatty-acid kinase homolog yqiU - *Bacillus subtilis*  
 C/Species: *Bacillus subtilis*  
 C/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 20-Jun-2000  
 R/Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter  
 S.; Bron, S.; Bouilliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chd  
 S.; Enrich, S.D.; Emerson, P.T.; Ehtan, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
 Nature 390, 249-256, 1997  
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall  
 tech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois  
 A/Authors: Laubex, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mause

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete  
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani  
 A/Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
 A/Reference number: A95580; MUID:98044033; PMID:9384377  
 A/Accession: C69562  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-363 <NUN>  
 A/Cross-references: GB:Z99116; GB:AL009126; NID:g2634723; PIDN:CAB14338.1; PID:g26348  
 A/Experimental source: strain 168  
 C/Genetics:  
 A/Genes: yqiU  
 C/Superfamily: acetate kinase

Query Match 33.8%; Score 46; DB 2; Length 363;  
 Best Local Similarity 71.4%; Pred. No. 50;  
 Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 12 AAEAVLKQVETIV 25  
 ||:||||:|  
 Db 293 AASAALKGEVAIV 306  
 ||:||||:|

RESULT 15  
 B95887  
 Probable aminotransferase protein [imported] - *Sinorhizobium meliloti* (strain 1021) m  
 C/Species: *Sinorhizobium meliloti*  
 C/Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C/Accession: B95887  
 R/Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A/Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing e  
 A/Reference number: A95842; MUID:21396508; PMID:11481431  
 A/Accession: B95887  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-461 <KUR>  
 A/Cross-references: GB:AL591985; PIDN:CAC48765.1; PID:gs15140238; GSPDB:GN00167  
 A/Experimental source: strain 1021, megaplasmid pSymB  
 R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubi  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau  
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,  
 A/Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A/Reference number: A96039; MUID:21368234; PMID:11474104  
 A/Contents: annotation  
 C/Genetics:  
 A/Genes: Smb20379  
 A/Genome: plasmid

Query Match 33.8%; Score 46; DB 2; Length 461;  
 Best Local Similarity 41.7%; Pred. No. 65;  
 Matches 10; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

Qy 4 FSTEMANKAAEAVLKQVETIVSF 27  
 ||:||||:|  
 Db 199 FATRMANNLSAILSEGPETIAAF 222  
 ||:||||:|

Search completed: November 13, 2003, 09:29:14  
 Job time: 6.81323 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 2.50584 Seconds  
(without alignments)  
525.472 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136

Sequence: 1 VYVFSTEMANKAEAVLKGQVETIVSFH 28

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	100.0	1426	1 BCL9_HUMAN	000512 homo sapien
2	98	72.1	1469	1 BCL9_DROME	Q961d9 orosophila
3	60	44.1	1205	1 HEMZ_CLOVO	Q529S5 clostridium
4	51	37.5	1217	1 SYV_FUGRU	P49696 fugu rubrip
5	46	33.8	330	1 QUTG_EMENI	P25416 emericeia
6	46	33.8	363	1 BUK_BACSU	P54532 bacillus su
7	46	33.8	2273	1 HFA1_YEAST	P32874 saccharomyc
8	45.5	33.5	224	1 SODM_CHAFE	Q96347 charybdis f
9	45.5	33.5	504	1 GUNN_ERWCA	Q52394 erwinia car
10	45.5	33.5	505	1 GUNV_ERWCA	Q53395 erwinia car
11	45.5	33.5	745	1 ZPT_MACRA	Q47096 erwinia car
12	45.5	33.5	745	1 ZPT_MACRA	O77726 macaca radi
13	45	33.1	490	1 PEM3_ARATH	Q966b9 arabidopsis
14	44.5	32.7	170	1 AIG2_ARATH	P54121 arabidopsis
15	44	32.4	318	1 AIG2_ARATH	Q58564 methanococc
16	44	32.4	340	1 MP6P_METJA	Q40614 caenorhabdi
17	44	32.4	484	1 SGAT_ECOLI	P39301 aeropyrum p
18	44	32.4	602	1 DNL1_AERPE	Q9ydl8 aeropyrum p
19	44	32.4	629	1 SYV_FVRPU	Q8u149 pyrococcus
20	44	32.4	980	1 SYV_SCHPO	Q75005 schizosacch
21	44	32.4	4427	1 PKSL_BACSU	Q05470 bacillus su
22	43	31.6	119	1 TH12_CASEL	Q17424 caenorhabdi
23	43	31.6	159	1 R154_SCHPO	Q9ubd1 schizosacch
24	43	31.6	291	1 E853_MYCPN	F75103 mycolobus
25	43	31.6	345	1 TRPD_SUITO	Q97127 scytolobus
26	43	31.6	537	1 ARSH_SCHPO	Q10269 schizosacch
27	43	31.6	575	1 YDEB_SCHPO	Q10499 schizosacch
28	43	31.6	594	1 SYV_FAT	Q04462 rattus norv
29	43	31.6	602	1 GUSL_HUMAN	Q9u132 homo sapien
30	43	31.6	609	1 NFRV_VIBPR	Q00971 vibrio prot
31	43	31.6	611	1 ENPA_VIBAN	P43147 vibrio angu
32	43	31.6	662	1 VGLF_CDVO	P12569 canine diet
33	43	31.6	895	1 SECA_CYACA	Q19911 cyanidiet c

## RESULT 1

ID	BCI9_HUMAN	STANDARD;	PRT;	1426 AA.
AC	000512;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).			
GN	BCL9.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID:9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Petal brain;			
RX	MEDLINE=98158621; PubMed=9490669;			
RA	Willis T.G., Zalcberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,			
RA	Jadavay D.M., Bastard C., Treleven J.G., Catovsky D., Silva W.L.M.,			
RA	Dyer M.J.S.;			
RT	"Molecular cloning of translocation t(1;14)(q21;q32) defines a novel			
RT	gene (BCL9) at chromosome 1q21";			
RU	Blood 9:11873-11881(1998).			
RN	[12]			
RP	FUNCTION.			
RX	MEDLINE=21952490; PubMed=11955446;			
RA	Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,			
RA	Murone M., Zuelig S., Basler K.;			
RA	"Wnt/wingless signaling requires BCL9/legless-mediated recruitment of			
RT	pygopus to the nuclear beta-catenin-TCF complex.";			
RL	Cell 109:47-60(2002).			
CC	!- FUNCTION: Involved in signal transduction through the wnt pathway.			
CC	!- SUBUNIT: Binds to beta-catenin (CTNBS1), PYGO1 and PYGO2.			
CC	!- SUBCELLULAR LOCATION: Nuclear (Probable).			
CC	!- TISSUE SPECIFICITY: Detected at low levels in thymus, prostate,			
CC	testis, ovary and small intestine, and at lower levels in spleen,			
CC	colon and blood.			
CC	!- DISEASE: Involved in a t(1;14)(q21;q32) chromosomal translocation			
CC	found in a patient with precursor B-cell acute lymphoblastic			
CC	leukemia (ALL). This translocation leaves the coding region			
CC	intact, but may have pathogenic effects due to alterations in the			
CC	expression level of BCL9. Several cases of translocations within			
CC	the 3' untranslated region of BCL9 have been found in B-cell			
CC	malignancies.			
CC	!- CAUTION: It is uncertain whether Met-1 or Met-27 is the initiator.			
CC	!- CAUTION: Ref.1 sequence differs from that shown due to a			
CC	frameshift in position 1391.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/</a> or			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			

## ALIGNMENTS



DR EMBL; Y13620; CAA73942.1; ALT\_FRAME.  
 DR Genew; HGNC:1008; BCL9.  
 DR MIM; 602597; .  
 DR GO; GO:0007048; P:oncogenesis; TAS.  
 DR Nuclear protein; Chromosomal translocation; Proto-oncogene;  
 KW Wnt signaling pathway.  
 FT DOMAIN 231 1378 PRO-RICH.  
 FT DOMAIN 347 377 CTNNB1-BINDING.  
 FT DOMAIN 331 335 POLY-PRO 1.  
 FT DOMAIN 514 517 POLY-PRO 2.  
 FT DOMAIN 900 903 POLY-ALA.  
 FT DOMAIN 970 973 POLY-PRO 3.  
 SQ SEQUENCE 1426 AA; 149314 MM; A240A487716B7F1B CRC64;  
 Query Match 100.0%; Score 136; DB 1; Length 1426;  
 Best Local Similarity 100.0%; Pred. No. 3.1e-12;  
 Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VVVFSTENANKAAEAVLKGVETTVSFH 28  
 Db 177 VVVFSTENANKAAEAVLKGVETTVSFH 204  
 RESULT 2  
 ID BCL9 DROME STANDARD; PRT; 1469 AA.  
 CC Q96159; Q9V4D2;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Bcl-9 homolog (Leigless protein).  
 GN BCL9 OR LGS OR CG2041  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Ephydroidea; Ephydroidea; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Branton R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter P.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 BA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Burtis K.C., Busan D.A., Butler J., Brokstein P., Brothier P.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,  
 RA Flooker A., Gong F., Gorrall J.H., Gu Z., Gelbart W.M., Glasser K.,  
 RA Glisler K., Gribble A.A., Hargraves J., Heine T.J., Hernandez J.R., Houck J.,  
 RA Harris N.L., Harvey D., Heiman T.J., Mei M.-H., Ibegwam C.,  
 RA Hostin D., Houston K.A., Howland T.J., Karpman G.H., Ketchum K.A.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry L., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Nusskern D.R., Pacle B.J.M.,  
 RA Palazzolo M., Nelson K.A., Nixon K., Nusskern D.R., Pacle B.J.M.,  
 RA Peltzer D.R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler P., Shen R.,  
 RA Shue B.C., Sidani-Kim S., Simpson M., Skupski M.P., Smith R.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Szekas R., Tector B., Turner R., Venter J., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yen K.F., Zehavi J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng K.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith R.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RN Science 287:2185-2195(2000).  
 RL SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley; TISSUE=Embryo;  
 MEDLINE=22426066; PubMed=12537569;  
 RC Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarini H., Krommiller B., Pacle B.J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celnik S.E.;  
 RA "A Drosophila full-length cDNA resource.";  
 RT Genom Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RL SEQUENCE OF 6-1469 FROM N.A., AND MUTAGENESIS OF GLY-514; LEU-534 AND  
 RP ILE-537.  
 RX MEDLINE=21952490; PubMed=11955446;  
 RA Kramps T., Peter O., Brunner E., Nellen D., Froesch B., Chatterjee S.,  
 RA Murone M., Zuellig S., Basler K.;  
 RT "Wnt/wingless signaling requires BCL9/legless-mediated recruitment of  
 RT PYGopus to the nuclear beta-catenin-TCF complex.";  
 CC Cell 109:47-60(2002).  
 CC -1- FUNCTION: Involved in signal transduction through the wnt pathway.  
 CC -1- SUBUNIT: Binds to ARM and PYGO.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically  
 CC throughout development.  
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation on  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by, non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/annouce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AF051651; AAK93075.1; AUT\_SEQ.  
 DR EMBL; AF457205; AAL91368.1; .  
 DR Flybase; FBgn0039907; lgs.  
 DR GO; GO:0005634; C:nucleus; IEP.  
 DR GO; GO:0030528; F:transcription regulator activity; IPI.  
 DR GO; GO:0030177; P:positive regulation of wnt receptor signal. .; IPI.  
 DR GO; GO:0007367; P:segment polarity determination; IMP.  
 KW Nuclear protein; Developmental protein; Segmentation polarity protein;  
 KW Wnt signaling pathway.  
 FT DOMAIN 511 555 ARM-BINDING.  
 FT DOMAIN 1134 1173 ASN-RICH.  
 FT DOMAIN 1340 1449 GLN-RICH.  
 FT DOMAIN 1162 1169 POLY-ASN.  
 FT MUTAGEN 514 514 G->E: IN ALLELE LGS-21L.  
 FT MUTAGEN 534 534 L->F: IN ALLELE LGS-17E; SEGMENT POLARITY  
 FT PHENOTYPE.  
 FT MUTAGEN 537 537 I->K: IN ALLELE LGS-17P.  
 SQ SEQUENCE 1469 AA; 153759 MM; 5672E01B7200ED08 CRC64;  
 Query Match 72.1%; Score 98; DB 1; Length 1469;  
 Best Local Similarity 72.1%; Pred. No. 1.6e-06;  
 Matches 16; Conservative 9; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 VVVFSTENANKAAEAVLKGVETTVSFH 28  
 Db 323 IFVFTQLANKAESVLSCQFTIIAH 350  
 RESULT 3  
 ID HEM2 CLOJO STANDARD; PRT; 205 AA.  
 AC Q59295;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Delta-aminolevulinic acid dehydratase (EC 4.2.1.24) (Porphobilinogen  
 DE synthase) (ALADH) (Fragment).  
 GN HEMW.  
 OS Clostridium josui.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium.  
 OC NCBI\_TaxID=1499;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PERM P-9684;  
 RX MEDLINE=95394829; PubMed=7665501;  
 RA Fujino E., Fujino T., Karita S., Sakka K., Ohmiya K.;  
 RT "Cloning and sequencing of some genes responsible for porphyrin  
 RT biosynthesis from the anaerobic bacterium Clostridium josui.";  
 RL J. Bacteriol. 177:5169-5175 (1995).  
 CC -!- CATALYTIC ACTIVITY: 2 5-aminolevulinate -> porphobilinogen + 2  
 CC H(2)O.  
 CC -!- COPACTOR: Zinc (By similarity).  
 CC -!- PATHWAY: Siroheme biosynthesis.  
 CC -!- SUBUNIT: Homotrimer (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE ALADH FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; D28503; BAA05863.1; -.  
 DR PIR; I40812; I40812.  
 DR HSSP; P45002; 1B4E.  
 DR InterPro; IPR001731; Alad\_dehydratase.  
 DR PRINTS; PF00490; ALAD; 1.  
 DR PRODOM; PD00144; DALDHVPTASE.  
 DR PROSITE; PD002304; ALAD\_DEHYDRATASE; 1.  
 KW Porphyrin biosynthesis; Lyase; Zinc.  
 FT DOMAIN 114 132 ZINC-BINDING (BY SIMILARITY).  
 FT NON\_TER 205 205  
 FT SEQUENCE 205 AA; 23172 MW; 88693DAEFD811445 CRC64;  
 SQ  
 Query Match 44.1%; Score 60; DB 1; Length 205;  
 Best Local Similarity 46.2%; Pred. NO. 0.1;  
 Matches 12; Conservative 5; Mismatches 9; Indels 0; Gaps 0;  
 QY 2 YVFSTEMANKAAEAVLKQVETIVSF 27  
 DB 51 YHFPDVGKAIKAAKADVKRSVLLF 76  
 RESULT 4  
 SYV FUGRU  
 ID - SYV FUGRU STANDARD; PRT; 1217 AA.  
 AC P49696;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (ValRS).  
 GN VARS1.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;  
 OC Tetraodonidae; Tetraodontidae; Takifugu.  
 RN NCBI\_TaxID=31033;  
 CC [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97396021; PubMed=9254008;  
 RX

RA Lim E.H., Corrochano L.M., Elgar G., Brenner S.;  
 RT "Genomic structure and sequence analysis of the valyl-tRNA synthetase  
 RT gene of the Japanese pufferfish, *Fugu rubripes*.";  
 RL DNA Seq. 7:141-151 (1997).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate  
 CC + L-valyl-tRNA(Val).  
 CC -!- SIMILARITY: Belongs to class-I aminoacyl-tRNA synthetase family.  
 CC -!- SIMILARITY: THE N-TERMINAL DOMAIN IS SIMILAR TO ELONGATION  
 CC FACTOR 1-GAMMA.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X01856; CAA62967.1; -.  
 DR HSSP; P96142; 1GAX.  
 DR InterPro; IPR004046; GST\_Cterm.  
 DR InterPro; IPR002300; tRNA-synt\_1a.  
 DR InterPro; IPR001412; tRNA-synt\_1.  
 DR InterPro; IPR002303; tRNA-synt\_val.  
 DR Pfam; PF00043; GST\_C; 1.  
 DR Pfam; PF00133; tRNA-synt\_1; 1.  
 DR PRINTS; PRO0986; TRNASYNTHVAL.  
 DR TIGRFAMs; TIGR00422; vals; 1.  
 DR PROSITE; PS00178; AA\_TRNA\_LIGASE\_1; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
 FT DOMAIN 1 7320 EF-1-GAMMA LIKE.  
 FT SITE 293 303 "HIGH" REGION.  
 FT SITE 809 813 "WMSKS" REGION.  
 FT BINDING 812 812 ATP (BY SIMILARITY).  
 FT SEQUENCE 1217 AA; 138218 MW; 5E08AF24B5C8A7A1 CRC64;  
 SQ  
 Query Match 37.5%; Score 51; DB 1; Length 1217;  
 Best Local Similarity 37.0%; Pred. NO. 14;  
 Matches 10; Conservative 8; Mismatches 9; Indels 0; Gaps 0;  
 QY 2 YVFSTEMANKAAEAVLKQVETIVSFH 28  
 DB 626 YVSCSDMGKQADAVREGRLKIPDH 652  
 RESULT 5  
 QUTG\_EMENI  
 ID QUTG\_EMENI STANDARD; PRT; 330 AA.  
 AC P25416;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE QUTG protein.  
 GN QUTG.  
 OS Emericella nidulans (Aspergillus nidulans).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; Emericella.  
 OC NCBI\_TaxID=162425;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89181521; PubMed=2976890;  
 RA Hawkins A.R., Lamb H.K., Smith M., Keyte J.W., Roberts C.F.;  
 RT "Molecular organisation of the quinac acid utilization (QUT) gene  
 RT cluster in *Aspergillus nidulans*.";  
 RL Mol. Gen. Genet. 214:224-231 (1988).  
 CC [2]  
 RP SIMILARITY TO INOSITOL MONOPHOSPHATASE.  
 RX MEDLINE=91080861; PubMed=2175387;  
 RA Lamb H.K., Hawkins A.R., Smith M., Harvey I.J., Brown J., Turner G.,  
 RA Roberts C.F.;  
 RT "Spatial and biological characterisation of the complete quinac acid  
 RT utilisation gene cluster in *Aspergillus nidulans*.";  
 RL Mol. Gen. Genet. 223:17-23 (1990).



Lin C.T., Lai Y.S., Kuo T.J., Chang T.C.:  
"Molecular cloning, expression, and characterization of a cDNA  
encoding Mn-superoxide dismutase from crab Charybdis feriatus";  
Submitted (Aug-1997) to The EMBL/GenBank/DBJ databases.  
- FUNCTION: Destroys radicals which are normally produced within the  
cells and which are toxic to biological systems  
- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
- COFACTOR: Manganese (By similarity).  
- COPOLYMER: Manganese (By similarity).  
- SUBUNIT: Homotrimer (By similarity).  
- SUBCELLULAR LOCATION: Mitochondrial matrix.  
- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE  
FAMILY.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see http://www.isb-sib.ch/announce/  
or send an email to license@isb-sib.ch).

-----  
EMBL; AF019411; AACD01640.1; -.  
DR HGSP; P04179; IABM.  
DR InterPro; IPR001189; SODismutase.  
DR Pfam; PF00081; sodfe; 1.  
DR Pfam; PF02777; sodfe; C.1.  
DR ProDom; PD000475; SODismutase; 1.  
DR PROSITE; PS00098; SOD\_MN; 1.  
DR Oxidoreductase; Mangastase; Mitochondrion; Transit peptide.  
KW TRANSIT 20 MITOXONDRIUM (BY SIMILARITY).  
FT CHAIN 21 224 SUPEROXIDE DISMUTASE [MN].  
FT METAL 46 46 MANGANESE (BY SIMILARITY).  
FT METAL 94 94 MANGANESE (BY SIMILARITY).  
FT METAL 177 177 MANGANESE (BY SIMILARITY).  
FT METAL 181 181 MANGANESE (BY SIMILARITY).  
SQ SEQUENCE 224 AA; 24527 MW; 107CF19382E9138A CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 224;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 13; Conservative 3; Mismatches 9; Indels 1; Gaps 1;

QY 2 YVFSTEWANKEASAVLVKQGVETVS 26  
|| : ||| ||| ||| |||  
Db 54 YVNNLNVAEEKLAAEKKGDVSTIIS 79

RESULT 9  
GUNN ERWCA STANDARD; PRT; 444 AA.  
AC QB9394;  
ID GUNN ERWCA  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Beta-glucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N)  
DE (Cellulase N).  
GN CELN.  
OS Erwinia carotovora.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Pectobacterium.  
OX NCBI\_TAXID=554;  
RN [1]  
RS SEQUENCE FROM N.A.  
RP STRAIN Atropspectica FCBR C18;  
RX MEDLINE=98299944; PubMed=963615;  
RA Olsen O., Thomsen K.K., Weber J., Duus J.O., Svendsen I., Wegener C.,  
von Wettstein D.;  
RT "Transplanting two unique beta-glucanase catalytic activities into  
one multicenzyme, which forms glucose.";  
RL Biotechnology 14:71-76 (1996).  
CC - CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic  
linkages in cellulose, lichenin and cereal beta-D-glucans.  
CC - SUBCELLULAR LOCATION: Secreted.  
CC - SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L39789; AAC37033.1; -
CC HSP; O85465; 1A3H.
CC InterPro; IPR001956; CBD 3.
CC InterPro; IPR001547; Glyco_hydro_5.
CC Pfam; PF00942; CBM 3; 1.
CC Pfam; PF00150; cellulase; 1.
CC ProDom; PD001947; CBD 3; 1.
CC ProSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
CC Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC SIGNAL 1 31 POTENTIAL.
CC CHAIN 32 504 ENDOGLUCANASE V1.
CC DOMAIN 32 334 CATALYTIC.
CC FT ACT SITE 168 168 PROTON DONOR (BY SIMILARITY).
CC FT ACT SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
CC FT ACT SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
CC SQ SEQUENCE 444 AA; 48300 MW; FAYEAI79004CB843 CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 444;
Best Local Similarity 36.7%; Pred. No. 33;
Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 2 YVFSTEMANKAAEAVLKQ---VETIVSFH 28
| : : ||| ||| : : |
Db 101 YISNPSLANKVKEAFAAAQSLGVYIIIDWH 130

RESULT 10
GUNW ERWCA STANDARD; PRT; 504 AA.
ID GUNW ERWCA STANDARD; PRT; 504 AA.
AC Q59195;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase VI precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V1)
DE (Cellulase VI).
GN ERWINIA CAROTOVORA.
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC3193;
RX MEDLINE=95231512; PubMed=7715600;
RA Mae A., Heikinhemo R., Palva E.T.;
RT "Structure and regulation of the Erwinia carotovora subspecies
RT carotovora SCC3193 cellulase gene celVI and the role of cellulase in
RT phytopathogenicity.";
RL Mol. Gen. Genet. 247:17-26(1995).
CC -! CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -! SUBCELLULAR LOCATION: Secreted.
CC -! SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X79241; CAAS55823.1; -
CC PIR; S54744; S54744.

```

```

DR HSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; cellulase; 1.
DR ProDom; PD001947; CBD 3; 1.
DR ProSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
DR Cellulose degradation; Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 31 POTENTIAL.
FT CHAIN 32 504 ENDOGLUCANASE V1.
FT DOMAIN 32 334 CATALYTIC.
FT FT ACT SITE 168 168 PROTON DONOR (BY SIMILARITY).
FT FT ACT SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
FT FT ACT SITE 256 256 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 504 AA; 54963 MW; 0D78CF74781565FA CRC64;

Query Match 33.5%; Score 45.5; DB 1; Length 504;
Best Local Similarity 36.7%; Pred. No. 38;
Matches 11; Conservative 5; Mismatches 11; Indels 3; Gaps 1;

QY 2 YVFSTEMANKAAEAVLKQ---VETIVSFH 28
| : : ||| ||| : : |
Db 101 YIANPSLANKVKEAFAAAQSLGVYIIIDWH 130

RESULT 11
GUNW ERWCA STANDARD; PRT; 505 AA.
ID GUNW ERWCA STANDARD; PRT; 505 AA.
AC Q47096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
GN CELV.
OS ERWINIA CAROTOVORA.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SCC1193;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmond G.P.C.;
RT "Molecular analysis of the major cellulase (CelV) of Erwinia
RT carotovora: evidence for an evolutionary 'mix-and-match' of enzyme
RT domains.";
RL Mol. Gen. Genet. 241:341-350(1993).
CC -! FUNCTION: Endoglucanase with some exoglucanase activity.
CC -! CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -! SUBCELLULAR LOCATION: Secreted.
CC -! MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature
CC optimum about 42 degrees Celsius.
CC -! SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X76000; CAAS3592.1; -
CC PIR; S39962; S39962.
DR HSP; O85465; 1A3H.
DR InterPro; IPR001956; CBD 3.
DR InterPro; IPR001547; Glyco_hydro_5.
DR Pfam; PF00942; CBM 3; 1.
DR Pfam; PF00150; cellulase; 1.

```

FT SIGNAL 1 38 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.  
FT CHAIN 39 745 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 717 736 POTENTIAL.  
FT TRANSMEM 737 745 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 370 537 ZP  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 105 105 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 122 122 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 223 223 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 269 269 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 310 310 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 400 400 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT SQ SEQUENCE 745 AA; 82710 MW; 731D9A4D3EE028 CRC64;  
Query Match 33.5%; Score 45.5; DB 1; Length 745;  
Best Local Similarity 29.7%; Pred. No. 57;  
Matches 11; Conservative 6; Mismatches 11; Indels 9; Gaps 1;  
QY 1 YVVFSTEMA-----NKAABAVLKQGVETVVSF 28  
DB 382 VYSQVOTAPDLQTLTVGNSSCPVFKAQSGLVRFH 418  
RESULT 13  
ID#1 PEM3 ARATH STANDARD; PRT: 490 AA.  
AC O9C6E9; O9C9V1;  
DC 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Putative phosphoethanolamine N-methyltransferase 3 [EC 2.1.1.103].  
GN WNT3 OR A1IG3600 OR F6D5.1 OR F25922.1.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosidia II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN=cv. Columbia;  
RC MEDLINE=21046719; PubMed=11130712;  
RA Theologis A., Ecker J.R., Falm C.J., Federspiel N.A., Kaul S.,  
RA White O., Alonso J., Altaf H., Araujo R., Bowman C.L., Brooks S.Y.,  
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,  
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,  
RA Dunn P., Egtu P., Feldbluth T.V., Feng J.-D., Fong B., Fujii C.Y.,  
RA Gill J.B., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,  
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,  
RA Klm C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,  
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,  
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,  
RA Milfischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,  
RA Pai G., Peterson J., Phan P.K., Rizzo M., Rooney T., Rowley D.I.,  
RA Sakao H., Salsberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,  
RA Sun H., Tallon L.J., Tanounga G., Toriumi M.J., Town C.D., Underhill T.,  
RA Unterbrink T., Van Aken S., Vaysberg M., Wysocka V.S., Walker M.,  
RA Wu D., Yu G., Fraser C.N., Venter J.C., Davis R.W.;  
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis  
thaliana.";  
RT Nature 408:816-820(2000).  
[2]  
RN CONCEPTUAL TRANSLATION.  
RP Schneider M.;  
RL Unpublished observations (MAY-2002).  
CC -1- FUNCTION: Catalyzes N-methylation of phosphoethanolamine,  
CC phosphomonoethanolamine and phosphodimethylethanolamine, the  
CC three methylation steps required to convert phosphoethanolamine to  
CC phosphocholine (By similarity).  
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + ethanolamine  
CC phosphate = S-adenosyl-L-homocysteine + N-methylethanolamine  
CC phosphate.  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

Search completed: November 13, 2003, 09:25:57  
Job time : 4.50584 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 13, 2003, 09:23:18 ; Search time 9.15175 Seconds  
(Without alignments)  
789.518 Million cell updates/sec

Title: US-09-915-543-15\_COPY\_177\_204

Perfect score: 136  
Sequence: 1 VYFSTEMAKAEAVLKGQETIVSFH 28

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mbc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55	40.4	1050	5 Q9U1Q4	Q9U1Q4 caenorhabdi
2	51	37.5	243	16 Q97JA0	Q97JA0 clostridium
3	51	37.5	1049	5 Q9V6L1	Q9V6L1 drosophila
4	51	37.5	1049	5 Q960E6	Q960E6 drosophila
5	50	36.8	459	10 Q82521	Q82521 capsicum ch
6	50	36.8	642	4 Q9H6R2	Q9H6R2 homo sapien
7	50	36.8	642	4 Q96GN2	Q96GN2 homo sapien
8	50	36.8	1098	4 Q96Q02	Q96Q02 homo sapien
9	49.5	36.4	350	2 Q8G3I9	Q8G3I9 lactobacill
10	49	36.0	141	10 Q9LFW7	Q9LFW7 arabidopsis
11	49	36.0	143	17 Q976C9	Q976C9 sulfolobus
12	49	36.0	187	16 Q8G3S3	Q8G3S3 bifidobacte
13	49	36.0	210	10 Q9FZU1	Q9FZU1 arabidopsi
14	49	36.0	233	17 Q96V19	Q96V19 sulfolobus
15	49	36.0	237	17 Q96XS5	Q96XS5 sulfolobus
16	49	36.0	268	16 Q8DFX2	Q8DFX2 vibrio vuln

17	49	36.0	458	16	Q83532	Q83532 treponema p
18	49	36.0	586	16	Q9KMS4	Q9KMS4 vibrio chol
19	49	36.0	1047	3	Q8TFZ1	Q8TFZ1 aspergillus
20	48.5	35.7	311	16	Q9PIC9	Q9PIC9 campylobact
21	48.5	35.7	688	5	Q96497	Q96497 trypanosoma
22	48	35.3	225	17	Q96250	Q96250 sulfolobus
23	48	35.3	235	17	Q976P7	Q976P7 sulfolobus
24	48	35.3	330	16	Q9K7G4	Q9K7G4 bacillus ha
25	48	35.3	609	10	Q9LV35	Q9LV35 arabidopsis
26	48	35.3	639	11	Q9BIN9	Q9BIN9 mus musculu
27	48	35.3	1034	3	Q8VTE9	Q8VTE9 candida gla
28	48	35.3	1659	4	Q8VWQ9	Q8VWQ9 homo sapien
29	47.5	34.9	444	16	Q8VXZ5	Q8VXZ5 clostridium
30	47	34.6	161	2	Q8G3E2	Q8G3E2 bartonella
31	47	34.6	359	17	Q8V4E0	Q8V4E0 arabidopsis
32	47	34.6	352	2	Q8V0S9	Q8V0S9 sulfolobus
33	47	34.6	399	16	Q8F1H0	Q8F1H0 streptococc
34	47	34.6	476	10	Q8RVV4	Q8RVV4 lycopersito
35	47	34.6	581	11	Q222Q3	Q222Q3 mus musculu
36	47	34.6	662	16	Q8Z7D3	Q8Z7D3 salmonella
37	47	34.6	732	16	Q8Z8A4	Q8Z8A4 shewanella
38	47	34.6	1072	3	Q8WZS5	Q8WZS5 neurospora
39	47	34.6	1261	5	Q8WMT3	Q8WMT3 dictyosteli
40	47	34.6	2388	5	Q8MYG5	Q8MYG5 dictyosteli
41	47	34.6	2473	12	Q71209	Q71209 grapevine 1
42	46.5	34.2	541	16	Q8EFD3	Q8EFD3 shewanella
43	46.5	34.2	614	5	Q22551	Q22551 caenorhabdi
44	46.5	34.2	771	5	Q19380	Q19380 caenorhabdi
45	46	33.8	320	5	Q81JX6	Q81JX6 plasmodium

## ALIGNMENTS

### RESULT 1

Q9U1Q4	PRELIMINARY;	PRT; 1050 AA.
ID	Q9U1Q4	
AC	Q9U1Q4;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Y87G2A.5 protein.	
GN	Y87G2A.5	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Pelodierinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	White S.;	
RL	Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=99069613; PubMed=9851916;	
RA	none;	
RT	"Genome sequence of the nematode C.elegans: A platform for	
RT	investigating biology."	
RL	Science 282:2012-2018(1998).	
DR	EMBL; AL110500; CAB60428.1; "	
DR	HSSP; P96142; 1GX.	
DR	WormPep; Y87G2A.5; CE24685.	
DR	InterPro; IPR002300; tRNA-synt 1a.	
DR	InterPro; IPR001412; tRNA-synt 1.	
DR	InterPro; IPR002303; tRNA-synt_val.	
DR	Pfam; PF00133; tRNA-synt 1; 1.	
DR	PRINTS; PRO0386; TRNASYNTHVAL.	
DR	TIGRFAMs; TIGR00422; vals; 1.	
DR	PROSITE; PS00178; AA TRNA LIGASE I; 1.	
DR	SEQUENCE 1050 AA, 118920 WK, F33DBS3587EAC057 CRC64;	
SQ		

Query Match 40.4%; Score 55; DB 5; Length 1050;  
Best Local Similarity 44.4%; Pred. No. 22;  
Matches 12; Conservative 2; Mismatches 13; Indels 0; Gaps 0;



QY 2 YVFSTEMANKAAEAAVLKQGVETIVSFH 28  
 Db 461 YVCAHMAEKAFAAANGDLOIPEFH 487

RESULT 2  
 Q97JAO PRELIMINARY: PRT; 243 AA.

AC Q97JAO  
 AC Q97JAO  
 DT 01-OCT-2001 (TEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)  
 DT 01-NAR-2002 (TEMBLrel. 20, Last annotation update)  
 DE Zn-dependent hydrolases, glyoxylase family.  
 GN CAC1386  
 OS Clostridium acetobutylicum.  
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;  
 OC Clostridium  
 OX NCBI\_TaxID=1488;  
 RN 11

SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;  
 RX MEDLINE=2135935; PubMed=11466286;  
 RA Noelling J., Bretton G., Omeichenko M.V., Makarova K.S., Zeng Q.,  
 RA Gibson R.L., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,  
 RA Tatusov R.N., Koonin E.V., Doucette-Stamm L., Soucaille P., Daly M.J.,  
 RA Bennett G.N., Koonin E.V., Smith D.R.;  
 RT "Genome sequence and comparative analysis of the solvent-producing  
 RT bacterium Clostridium acetobutylicum.";  
 RL J. Bacteriol. 183:4823-4838(2001).  
 DR EMBL; AE007650; AAK79354.1; -  
 DR InterPro; IPR001279; Blackmase-like.  
 DR Pfam; PF00753; lactamase B; 1.  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 243 AA; 26880 MW; 0F9F6A3EADBEBCD CRC64;

Query Match 37.5%; Score 51; DB 16; Length 243;  
 Best Local Similarity 25.0%; Pred. No. 18;  
 Matches 7; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 1 YVFSTEMANKAAEAAVLKQGVETIVSFH 28  
 Db 199 LPFDNSLKSLEKTKYDITVICH 226

RESULT 3  
 Q9V6L1 PRELIMINARY: PRT; 1049 AA.

AC Q9V6L1  
 AC Q9V6L1  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DT 01-NAR-2003 (TEMBLrel. 23, Last annotation update)  
 DE CG4062 protein.  
 GN AATS-VAL OR CG4062.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN 11

SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.C., Wortman J.K., Fendley M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Van K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abiri J.F., Agbayani A., An H.-J., Andrews-Fianknoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,  
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferraz C., Fleischmann W.,  
 RA Fosler K., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Thegawam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Lin X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 RN 12

SEQUENCE FROM N.A.  
 RA Celnik S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H. Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phoumenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN 13

SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Trudy J.L., Bergman C., Berman B., Carlson J.W., Celnik S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN 14

SEQUENCE FROM N.A.  
 RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN 15

SEQUENCE FROM N.A.  
 RA FlyBase;  
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB003819; AAM68598.1; -  
 DR HSSP; P96142; 1GAX.  
 DR FlyBase; FBgn0027079; Aats-val.  
 DR InterPro; IPR002300; tRNA-synt\_1a.  
 DR InterPro; IPR001412; tRNA-synt\_1.  
 DR InterPro; IPR002303; tRNA-synt\_val.  
 DR Pfam; PF00133; tRNA-synt\_1; 1.  
 DR PRINTS; PRO0986; TRNASYNTHAL.  
 DR TIGRFBAS; TIGR00422; vals; 1.

Accession	Gene	Accession	Gene
RA	Aluru M., Curry J., O'Connell M.,	RA	Scarsberg R.,
RT	"Nucleotide Sequence of a Probable	RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RT	No. AF085149) from Habanero Chile. (PGR98-182).",	DR	EMBL; BC009355; AAH09355.1; -.



**RA** Pridmore R.D., Arigoni F.;  
The genome sequence of *Bifidobacterium longum* reflects its adaptation  
**RT** to the human gastrointestinal tract.";  
**RL** Proc. Natl Acad Sci U.S.A. 99:14422-14427(2002).  
**RK** EMBL; AF014802; JN25468.1.  
**KW** Glycosyltransferase, Transferase, Complete Proteome.  
**DQ** SEQUENCE 187 AA; 20617 MW; 5697C3C12727D81 CRC64;

Query Match 36.0%; Score 49; DB 16; Length 187;  
Best Local Similarity 45.2%; Pred NC; 26;  
Matches 14; Conservative 3; Mismatches 4; Indels 10; Gaps 1;

**OY** 7 EMANKAAR-----NLKGQVETIVSF 27  
||||| :||| ||||| :|||:  
**Db** 26 EMAALASDYDKNPVLVALVGAVNTLVAF 56  
::: :::::

**RESUL 13**

**ID** Q9FZLL PRELMINARY; PRT; 210 AA.

**AC** Q9FZLL;

**DT** 01-MAR-2001 (TEMBUREl. 16, Created)

**EJ** 01-MAR-2001 (TEMBUREl. 16, Last sequence update)

**ST** 01-OCT-2001 (TEMBUREl. 18, Last annotation update)

**DE** F17L21.2.

**OS** Arabidopsis thaliana (Mouse-ear cress).

**OC** Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

**CC** Spermatopsids II; Magnoliophyta; eudicotyledons; core eucots; Rosidae;

**OX** NCBI\_TaxID=3702;

**[1]**

**RP** SEQUENCE FROM N.A.

**RA** Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C., Shinn P., Altafi H., Bei Q., Chin C., Chiou J., Choi E., Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharly N., Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A., Torumi M., Vaysberg M., Yu G., Fedetspiel N.A., Theologis A., Eckher J.R.;

**RT** "Genomic sequence for Arabidopsis thaliana BAC F17L21 from Chromosome I";

**RN** Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.

**[2]**

**RP** SEQUENCE FROM N.A.

**RA** Eckher J.R.;

**RL** Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.

**[3]**

**RP** SEQUENCE FROM N.A.

**RA** Eckher J.R.;

**RL** Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.

**[4]**

**RP** SEQUENCE FROM N.A.

**RA** Eckher J.R.;

**RL** Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.

**[5]**

**RP** SEQUENCE FROM N.A.

**RA** Eckher J.R.;

**RL** Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

**[6]**

**RP** SEQUENCE FROM N.A.

**RA** Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E., Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B., Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharly N., Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A., Torumi M., Vaysberg M., Yu G., Davis R., Federspigel N., Theologis A., Eckher J.;

**RL** Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.

**DR** EMBL; AC004557; XAF9722.1 - "

**RF** InterPro: IPRO03822; PAH.

**PfAM** PF02671; PAH; 1.

**SQ** SEQUENCE 210 AA; 24697 MW; FCDF8130CD75700AO CRC64;

```

Query Match          36.0%; Score 49; DB 10; Length 210;
Best Local Similarity 41.7%; Pred. No. 30;
Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      4  ESTEMANKAAEAVLKGQVETVSP 27
      |||||:|:|:|:|:|:|:|:|:|:|
Db      181 WSPFSTNQADRKLGELENNVIF 204

RESULT 14
Q96YA9
ID      Q96YA9          PRELIMINARY;      PRT;      233 AA.
AC      Q96YA9;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DD      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      Hypothetical protein ST2259.

```

Query Match 36.0%; Score 49; DB 17; Length 233;  
Best Local Similarity 39.3%; Pred. No. 34;  
Matches 11; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

```

RESULT 15
Q96X35 PRELIMINARY; PRT; 237 AA.
ID Q96X55
AC Q96X55
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DD 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DD DE Hypothetical protein Stil165.
DN Stil165 OR S0857.
GN Sulfolobus tokodaii.
OS Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
NCBI_TaxID=111955;
[1]
RN SEQUENCE FROM N.A.
RP RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RR Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RR Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RR Oshima T., Kikuchi H.;
RN "Complete genome sequence of an aerobic thermacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain".

```

